The Journal of Chemical Physics

Second-order MCSCF optimization revisited. I. Improved algorithms for fast and robust second-order CASSCF convergence

Cite as: J. Chem. Phys. **150**, 194106 (2019); https://doi.org/10.1063/1.5094644
Submitted: 04 March 2019 . Accepted: 25 April 2019 . Published Online: 20 May 2019

David A. Kreplin, Peter J. Knowles, and Hans-Joachim Werner

AIP Publishing

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Second-order MCSCF optimization revisited. I. Improved algorithms for fast and robust second-order CASSCF convergence

Cite as: J. Chem. Phys. 150, 194106 (2019); doi: 10.1063/1.5094644
Submitted: 4 March 2019 • Accepted: 25 April 2019 •
Published Online: 20 May 2019

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

AFFILIATIONS

[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] Electronic mail: kreplin@theochem.uni-stuttgart.de
[b] Electronic mail: KnowlesPJ@Cardiff.ac.uk
[c] Electronic mail: werner@theochem.uni-stuttgart.de

ABSTRACT

A new improved implementation of the second-order multiconfiguration self-consistent field optimization method of Werner and Knowles [J. Chem. Phys. **82**, 5053 (1985)] is presented. It differs from the original method by more stable and efficient algorithms for minimizing the second-order energy approximation in the so-called microiterations. Conventionally, this proceeds by alternating optimizations of the orbitals and configuration (CI) coefficients and is linearly convergent. The most difficult part is the orbital optimization, which requires solving a system of nonlinear equations that are often strongly coupled. We present a much improved algorithm for solving this problem, using an iterative subspace method that includes part of the orbital Hessian explicitly, and discuss different strategies for performing the uncoupled optimization in a most efficient manner. Second, we present a new solver in which the orbital-CI coupling is treated explicitly. This leads to quadratic convergence of the microiterations but requires many additional evaluations of reduced (transition) density matrices. In difficult optimization problems with a strong coupling of the orbitals and CI coefficients, it leads to much improved convergence of both the macroiterations and the microiterations. Third, the orbital-CI coupling is treated approximately using a quasi-Newton approach with Broyden–Fletcher–Goldfarb–Shanno updates of the orbital Hessian. It is demonstrated that this converges almost as well as the explicitly coupled method but avoids the additional effort for computing many transition density matrices. The performance of the three methods is compared for a set of 21 aromatic molecules, an Fe(II)-porphine transition metal complex, as well as for the $[Cu_2O_2(NH_3)_6]^{2+}$, $FeCl_3$, $Co_2(CO)_6C_2H_2$, and Al_4O_2 complexes. In all cases, faster and more stable convergence than with the original implementation is achieved.

I. INTRODUCTION

The multiconfiguration self-consistent field (MCSCF) method[1–45] is widely applied in quantum chemistry to handle strong electron correlation effects, where the wavefunction is dominated by more than one configuration. It is important for computing global potential energy surfaces and excited states or in cases where two or more states become nearly degenerate. MCSCF wavefunctions are also the starting point for subsequent treatments of dynamical correlation effects, using either multireference perturbation theory (MRPT),[46–52] multireference configuration interaction (MRCI),[53–61] or multireference coupled-cluster (MRCC) methods.[62–65] In MCSCF methods, the orbitals and configuration expansion coefficients (denoted CI coefficients) are optimized simultaneously by minimization of the energy. If excited states are considered, the weighted energy average of several states is usually optimized, yielding a single set of orbitals to describe all states (SA-MCSCF).[5,9,18] In general, the orbitals and CI coefficients depend on each other, leading to a difficult nonlinear optimization problem, which must be solved iteratively and which typically contains near redundancy between some combinations of orbital rotations and CI coefficients.

Apart from this, the proper choice of the orbitals and configurations to be included in an MCSCF wavefunction is often difficult. The most widely used ansatz is to include all possible configurations that can be generated within a preselected active space of orbitals, leading to the complete active space self-consistent field (CASSCF) method.[10,15,66,67] However, despite this formally simple ansatz, much experience and many trials are often necessary until a reasonable active space is found. New tools[68–71] have recently been proposed for simplifying the construction and selection of the active space and driving the CASSCF method closer to a black-box application, but this problem is still not yet generally solved.

In CASSCF methods, the orbitals can be classified into inactive orbitals, which are doubly occupied in all configurations, active orbitals with varying occupations, and virtual (or external) orbitals which are unoccupied. The optimization of CASSCF wavefunctions is somewhat simplified by the fact that orbital rotations within the active space are redundant; i.e., they can be described solely by changes in the CI-coefficients. However, a commonly encountered problem is that active and inactive or virtual orbitals may swap during the optimization. Furthermore, the energy may depend weakly on rotations between inactive orbitals and nearly doubly occupied active orbitals or on rotations between active orbitals with small occupation numbers and virtual orbitals. All this can make CASSCF optimizations difficult and slowly convergent.

Another problem is the factorial increase in the number of configuration state functions (CSFs) or Slater determinants (SDs) with the number of active orbitals. Depending on the number of active electrons, the spin quantum numbers, and the spatial symmetry, this limits the number of active orbitals that can be handled by conventional approaches to 14–18. Recently, FCI (full configuration interaction) calculations with up to 22 electrons in 22 active orbitals [CAS(22,22)] have been reported by employing massive parallelization.[72] In the last decade, several new approximate FCI methods have been introduced, which allow the treatment of considerably larger active spaces. They are based on stochastic approximations, for example, the full configuration interaction quantum Monte Carlo (FCIQMC) framework[73–75] or the heat-bath CI.[76,77] Other deterministic approaches include incremental expansions[78] or the density matrix renormalization group (DMRG) algorithm, which employs a CI parameterization with a lower degree of freedom.[79] The DMRG algorithm has also been successfully integrated into CASSCF methods in different ways.[44,45,80–83] However, with large active spaces such optimizations are extremely expensive. Furthermore, due to the approximations in the FCI treatment the redundancy between the active orbital rotations and CI-coefficients is lifted, which can slow down the convergence.[45]

The first developments of MCSCF methods date back nearly half a century (see Ref. 3 for a bibliography of early work), and since then, many different numerical approaches have been applied to the problem. In general, one can distinguish first-order and second-order optimization algorithms. First-order methods[9–15,42,43] mostly decouple the CI and orbital optimizations. In each iteration, first the CI eigenvalue problem is (approximately) solved with a given set of orbitals. Using the new CI coefficients, the gradient with respect to orbital rotations is computed and employed to update the orbitals. For example, this can be done using the so-called super-CI method,[4,7–11] which is based on the generalized Brillouin theorem.[2] The orbital gradient depends on the one- and two-particle reduced density matrices (RDMs), which in turn depend on the CI-coefficients. With the updated orbitals, the integrals are transformed and then the process is repeated. In the following, we will denote the alternating optimization of orbitals and CI coefficients as "uncoupled" optimization. In simple cases, this converges reasonably well, but in more difficult applications, hundreds of iterations may be needed to reach convergence.

In second-order optimization methods,[13–40,44,45] the second derivatives of the energy with respect to the orbital rotations and the CI-coefficients are also explicitly or implicitly included. The general way is to define an approximate energy functional that is accurate to second order in small changes in the orbitals and CI coefficients. In each "macroiteration", the approximate functional is minimized iteratively. These iterations are denoted "microiterations". They can be carried out with a single set of integrals, which are computed with the initial orbitals. After the minimum has been found, new orbitals are obtained, and then new integrals for the next macroiteration are computed using a 4-index integral transformation. Near the final solution, second-order methods achieve quadratic convergence of the energies obtained in successive macroiterations.

Expanding the energy up to second-order in a unique set of orbital rotation parameters $\{R_{ri}\}$ and parameters $\{\Delta c_I\}$ describing the changes in the CI-vectors and minimizing the resulting energy functional leads to a system of linear equations, which can be solved for $\mathbf{R}$ and $\Delta\mathbf{c}$ [Newton-Raphson (NR) method]. Unfortunately, despite formal quadratic convergence, this method has a small radius of convergence since the Hessian matrix of second derivatives frequently has many negative eigenvalues in the initial iterations. Level shifting and step restriction approximations[23,32,33] can be used to modify the Hessian to a positive definite form. This can be conveniently achieved with a damped generalization of the so-called Augmented Hessian (AH) method,[84] in which an eigenvalue problem, rather than a linear equation set, has to be solved. Convergence toward an energy minimum can always be enforced by sufficiently large level shifts and/or damping parameters, but the convergence may then become quite slow.

A principal reason for these problems is that the energy is a periodic function of the individual orbital rotations, and this is only poorly approximated by a second-order expansion in $\mathbf{R}$. Werner and Meyer[17,18] therefore proposed to expand the energy up to second order in $\mathbf{T} = \mathbf{U} - \mathbf{1}$ rather than in $\mathbf{R}$, where $\mathbf{U} \equiv \mathbf{U}(\mathbf{R})$ is an orthogonal orbital transformation matrix, and showed that this leads to a dramatic improvement of the speed and radius of convergence. Minimizing the second-order energy expansion $E^{(2)}(\mathbf{T})$ requires exactly the same integrals as minimizing $E^{(2)}(\mathbf{R})$, but leads to a system of nonlinear equations, which are more difficult to solve than the NR or AH equations. The method has further been improved and extended to large CASSCF cases by Werner and Knowles[19,20] and will in the following be denoted as the WMK method. It is part of the MOLPRO quantum chemistry package[85,86] and has in the last 30 years been used in a wide variety of successful applications. Recently, closely related methods, which also use expansions in $\mathbf{T}$, have been developed by other authors as well.[44,45] Other recent

developments include atomic orbital (AO)-driven schemes with first or second-order convergence.[43,44] Even though these AO-driven schemes are formally more expensive than methods based on explicit integral transformations, they are suitable for deployment on graphical processing unit (GPU) based computers,[43,87–90] and exploiting the high computational power of GPUs and the sparsity of the integrals in the AO basis has made it possible to carry out calculations for large systems with very large basis sets.

The main difficulty in the WMK method is to converge the microiterations efficiently and reliably. On the one hand, the orbital optimization requires solving a system of nonlinear equations, which are often rather strongly coupled. On the other hand, the coupling between the orbitals and CI coefficients can also be strong. The purpose of the current paper is to describe improvements in the algorithms used to solve the microiterations in the WMK method. This includes several aspects:

(i) Better convergence of the microiterations for minimizing $E^{(2)}(\mathbf{T}, \mathbf{c})$ for fixed CI coefficients $\mathbf{c}$. This includes a subspace (P-space) optimization method for solving large AH equations or large CI eigenvalue problems for (nearly) degenerate states.
(ii) An improvement of the AH method in the case that the lowest eigenvectors become (nearly) degenerate. In previous implementations, this problem could lead to nondeterministic convergence behavior of the microiterations.
(iii) The choice of an optimal strategy for alternating orbital and CI optimization steps.
(iv) Improvements in the convergence and efficiency through explicit inclusion of the orbital-CI coupling, leading to quadratic convergence of $E^{(2)}(\mathbf{T}, \mathbf{c})$.
(v) An approximate inclusion of the orbital-CI coupling using a quasi-Newton (QN) approximation.

To the best of our knowledge, a quadratically convergent optimization of the second-order energy approximation $E^{(2)}(\mathbf{T}, \mathbf{c})$ has not been described and implemented before.

Applying the WMK method or other second-order MCSCF optimization approaches for large molecules with very many inactive orbitals may become extremely expensive, since the computational effort for the integral transformation scales with the fifth power [$\mathcal{O}(\mathcal{N}^5)$] of the molecular size $\mathcal{N}$, and the storage of the transformed integrals and other intermediates scales as $\mathcal{O}(\mathcal{N}^4)$. This problem can be alleviated in two ways: One possibility is to take a large part of the inactive orbitals from a preceding Hartree-Fock calculation and keep them frozen. This works well if the active orbitals are localized in some part of the molecule (e.g., at a transition metal and its environment) and if a localized representation of the inactive orbitals is used.[91] Assuming that this leads to the number of optimized orbitals being independent of the molecular size, the central processing unit (CPU) time scaling is reduced to $\mathcal{O}(\mathcal{N}^4)$ (with conventional integral transformation) or $\mathcal{O}(\mathcal{N}^3)$ [with density-fitting (DF) or Cholesky decomposition (CD) approximations]. Storage then scales only as $\mathcal{O}(\mathcal{N}^2)$. Another possibility is to fully optimize the inactive orbitals using a first-order algorithm, coupled to a second-order optimization of the active orbitals, leading to the possibility to maintain the fast and robust convergence of the second-order optimization at much reduced cost [scaling as $\mathcal{O}(\mathcal{N}^3)$ if DF is used]. This new hybrid method will be described in Paper II.[130]

The paper is organized as follows. In Sec. II, we briefly review the WMK method and present the working equations that are relevant in this work. Section III discusses improvements of the conventional algorithm, in which the orbitals and CI coefficients are optimized in an alternating way. In Sec. IV, explicit inclusion of the coupling of orbital rotations and CI coefficients is discussed, and in Sec. V, this coupling is approximated by a QN scheme. In Sec. VI, we present benchmark calculations to investigate the influence of the coupling on the convergence and discuss the best overall optimization strategy.

II. REVIEW OF THE WERNER-KNOWLES SOLVER

We consider the MCSCF wavefunction for state n

$$|\Psi_n\rangle = \sum_I |\Phi_I\rangle c_I^n, \tag{1}$$

where $|\Phi_I\rangle$ represents either a CSF or a SD, and c_I^n are the CI coefficients. The configurations are constructed from the molecular orbitals (MOs) $|k)$. Here and in the following (unless otherwise noted), the indices k, l will refer to any occupied orbitals (i.e., those that appear in at least one $|\Phi_I\rangle$), indices a, b will refer to virtual orbitals (unoccupied in all configurations), and r, s will refer to any orbitals. All orbitals are assumed to be real. The occupied orbital space can be divided into inactive (closed-shell) and active subspaces, which are labeled with indices i, j and t, u, v, w, respectively. The orbitals are varied by an orthogonal transformation $\mathbf{U}$,

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

which is parameterized as

$$\mathbf{U} = \exp(\mathbf{R}) = \mathbf{1} + \mathbf{R} + \frac{1}{2}\mathbf{R}^2 + \cdots = \mathbf{1} + \mathbf{T}. \tag{3}$$

The antisymmetric matrix $\mathbf{R} = -\mathbf{R}^\dagger$ contains the independent orbital rotation parameters R_{rk} (redundant rotation parameters within the inactive, active, and virtual orbital subspaces are set to zero). For a given set of orbitals and CI-coefficients, the state energy can be written as

$$E_n = \sum_{tu} F_{tu}^c D_{tu}^n + \frac{1}{2}\sum_{tuvw} (tu|vw) D_{tu,vw}^n + E_c + E_{nuc}, \tag{4}$$

where the closed-shell Fock matrix F_{rs}^c and the closed-shell energy E_c are defined as

$$F_{rs}^c = h_{rs} + \sum_j [2(rs|jj) - (rj|js)], \tag{5}$$

$$E_c = \sum_j (h_{jj} + F_{jj}^c). \tag{6}$$

h_{rs} are the one-electron integrals and $(rs|tu)$ are the two-electron repulsion integrals (in Mulliken notation). E_{nuc} is the nuclear repulsion energy. For later use, we define tensors of two-electron integrals with at least 2 occupied indices $J_{rs}^{kl} = (rs|kl)$ and $K_{rs}^{kl} = (rk|sl)$. These are considered as matrices $\mathbf{J}^{kl}$, $\mathbf{K}^{kl}$ with indices r, s.

D^n_{tu} and $D^n_{tu,vw}$ are the 1-particle and 2-particle reduced density matrices (1-RDM and 2-RDM), projected onto the active orbital space, for state n, respectively, which depend solely on the CI-coefficients,

$$D^n_{tu} = \sum_{I,J} c^n_I c^n_J D^{IJ}_{tu} \quad \text{and} \quad D^n_{tu,vw} = \sum_{I,J} c^n_I c^n_J D^{IJ}_{tu,vw}. \tag{7}$$

The quantities $D^{IJ}_{tu} = \langle \Phi_I | \hat{E}_{tu} | \Phi_J \rangle$ and $D^{IJ}_{tu,vw} = \langle \Phi_I | \hat{E}_{tu,vw} | \Phi_J \rangle$ are denoted coupling coefficients. They are independent of the orbitals and CI-coefficients and depend only on the structure of the CSFs. The 2-particle density matrices $D^n_{tu,vw}$ are not symmetric with respect to permutation of t, u or v, w. However, since in the energy expression they are contracted across at least t, u or v, w with the symmetric integrals $(tu|vw)$, they can be symmetrized as

$$\tilde{D}^n_{tu,vw} = \frac{1}{2}(D^n_{tu,vw} + D^n_{ut,vw}) \tag{8}$$

so that the permutation symmetries of the indices t, u, v, w are the same as for the integrals. In the following, the bar over D will be omitted and the density matrices will always be assumed to be symmetrized.

In state-averaged MCSCF, the energy average $E_{\text{av}} = \sum_n W_n E_n$ is optimized, where W_n are the weights, $\sum_n W_n = 1$. In this case, E_{av} can be constructed using Eq. (4) with state averaged RDMs $D^{\text{av}}_{tu} = \sum_n W_n D^n_{tu}$ and $D^{\text{av}}_{tu,vw} = \sum_n W_n D^n_{tu,vw}$. The remaining formalism in this section is the same for single-state MCSCF and SA-MCSCF, and we will therefore omit the state labels in the following. E_0 will denote the energy obtained with the initial orbitals and CI coefficients in a macroiteration.

In the WMK method, the energy is expanded up to second order in the orbital changes, i.e., up to second order in $\mathbf{T}$ and $\mathbf{c}$. This takes care of the periodicity of the energy with respect to individual orbital rotations and has a much larger radius of convergence than expanding the energy up to second order in the parameters R_{rk}. For fixed CI-coefficients, the second-order expansion of the energy in $\mathbf{T}$ can be written as

$$E^{(2)}(\mathbf{T}) = E_0 + \sum_{r,k} T_{rk}\Big[2A_{rk} + \sum_{s,l} G^{kl}_{rs} T_{sl}\Big]. \tag{9}$$

The matrices A_{rk} and G^{kl}_{rs} are defined as[19–21]

$$A_{ri} = 2F_{ri}, \tag{10}$$

$$A_{ru} = \sum_t F^c_{rt} D_{tu} + \sum_{t,v,w} (rt|vw) D_{tu,vw}, \tag{11}$$

$$A_{ra} = 0, \tag{12}$$

$$G^{ij}_{rs} = 2\Big[F_{rs}\delta_{ij} + L^{ij}_{rs}\Big], \tag{13}$$

$$G^{tj}_{rs} = \sum_v D_{tv} L^{vj}_{rs} = G^{jt}_{sr}, \tag{14}$$

$$G^{tu}_{rs} = F^c_{rs} D_{tu} + \sum_{v,w} [J^{vw}_{rs} D_{tu,vw} + 2K^{vw}_{rs} D_{tv,uw}], \tag{15}$$

where

$$L^{kj}_{rs} = 4\,K^{kj}_{rs} - K^{kj}_{sr} - J^{kj}_{rs}, \tag{16}$$

$$F_{rs} = F^c_{rs} + \sum_{t,u} D_{tu}\Big[J^{tu}_{rs} - \frac{1}{2}K^{tu}_{rs}\Big]. \tag{17}$$

Minimization of the energy with respect to $\mathbf{T}$, subject to the orthonormality condition of the orbitals, leads to the condition

$$\mathbf{U}^\dagger \mathbf{B} - \mathbf{B}^\dagger \mathbf{U} = 0 \tag{18}$$

with

$$B_{rk} = A_{rk} + \sum_{sl} G^{kl}_{rs} T_{sl}, \quad B_{ra} = 0. \tag{19}$$

For the fully optimized orbitals, the solution of Eq. (18) must be $\mathbf{U} = \mathbf{1}$, and thus, the variational conditions are

$$A_{ak} = 0 \quad \text{and} \quad A_{kl} - A_{lk} = 0. \tag{20}$$

For a given $\mathbf{U} = \mathbf{1} + \mathbf{T}$, the Hamiltonian can be expanded as

$$H^{(2)}_{IJ} = \sum_{tu} F^{c\,(2)}_{tu} D^{IJ}_{tu} + \frac{1}{2}\sum_{tuvw} (tu|vw)^{(2)} D^{IJ}_{tu,vw} + \delta_{IJ} E^{(2)}_c \tag{21}$$

with the second-order expansions of the integrals and closed-shell energy

$$E^{(2)}_c = E_c + 4\sum_i [\mathbf{F}^c \mathbf{T}]_{ii} + 2\sum_{ij} [\mathbf{T}^\dagger (\mathbf{F}^c \delta_{ij} + \mathbf{L}^{ij})\mathbf{T}]_{ij}, \tag{22}$$

$$F^{c\,(2)}_{tu} = (\mathbf{U}^\dagger \mathbf{F}^c \mathbf{U})_{tu} + \sum_i \Big[2\big(\mathbf{U}^\dagger \mathbf{J}^{tu} \mathbf{U} - \mathbf{J}^{tu}\big)_{ii} - \big(\mathbf{U}^\dagger \mathbf{K}^{tu} \mathbf{U} - \mathbf{K}^{tu}\big)_{ii} + (\mathbf{T}^\dagger \mathbf{L}^{ui} \mathbf{T})_{ti} + (\mathbf{T}^\dagger \mathbf{L}^{ti} \mathbf{T})_{ui}\Big], \tag{23}$$

$$(tu|vw)^{(2)} = -(tu|vw) + (\mathbf{U}^\dagger \mathbf{J}^{vw} \mathbf{U})_{tu} + (\mathbf{U}^\dagger \mathbf{J}^{tu} \mathbf{U})_{vw} + (1 + \tau_{tu})(1 + \tau_{vw})(\mathbf{T}^\dagger \mathbf{K}^{tv} \mathbf{T})_{uw}. \tag{24}$$

The second-order energy can then be written in the alternative form

$$E^{(2)}(\mathbf{T}) = \sum_{I,J} c_I c_J H^{(2)}_{IJ}, \quad \sum_I c_I^2 = 1, \tag{25}$$

which is equivalent to Eq. (9). Thus, the variational conditions for the CI-coefficients are

$$\Big[\mathbf{H}^{(2)} - E^{(2)}_n \mathbf{1}\Big]\mathbf{c}^n = \mathbf{0}. \tag{26}$$

For $\mathbf{U} = \mathbf{1}$, this reduces to the standard CI eigenvalue problem. In each macroiteration, Eqs. (18) and (26) have to be satisfied simultaneously. Solving these equations constitutes the microiterations. In the beginning of each macroiteration, the integral matrices $\mathbf{J}^{kl}$ and $\mathbf{K}^{kl}$ must be computed. The integral transformation scales with $O(N^5)$ and is in many cases the dominating step per macroiteration.

The energy approximation $E^{(2)}(\mathbf{T})$ is very accurate for the occupied-virtual rotations. However, it fails to represent the correct behavior between inactive and nearly doubly occupied active orbitals.[19,21,92] For these internal rotations, the steps are often underestimated. To improve the convergence of the rotations between active and inactive orbitals, each macroiteration is started with an additional optimization of the CI coefficients and the nonredundant internal orbital rotations, using the exact energy expression. During the internal optimization, only the integrals $(kl|mn)$ (k, l, m, n being occupied orbitals) need to be transformed, which scales as n^5_{occ} and is cheap. At the end of the internal optimization, the full J^{kl}_{rs} and K^{kl}_{rs} matrices transformed as

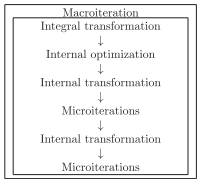

FIG. 1. Optimization scheme of a single macroiteration.

$$J^{k'l'}_{r's'} = \sum_r U_{rr'} \sum_s U_{ss'} \sum_k U_{kk'} \sum_l U_{ll'} J^{kl}_{rs}, \tag{27}$$

but since only the occupied-occupied block of **U** is used in the transformation of r, s (i.e., $U_{ra} = U_{ar} = \delta_{ar}$), these transformations only scale as $n_{\mathrm{orb}} n^4_{\mathrm{occ}}$ and are also cheap. A further improvement can be achieved by repeating this internal integral transformation after convergence of the microiterations and subsequently repeating the microiterations with the transformed integrals. In our new implementation, the internal optimization is taken from the WMK method and (apart from some technical improvements) kept unchanged.

An overview of one single macroiteration is shown in Fig. 1. Provided that the microiterations converge so that both variational conditions (18) and (26) are satisfied simultaneously, the macroiterations converge rapidly for reasonably chosen starting guesses and only very few full integral transformations are necessary.

In Secs. III–V, we describe three different algorithms for solving the microiterations. First, we review the commonly used alternating optimization, which decouples the optimization of the CI orbitals and orbitals. Various improvements of this algorithm are described. Next, we present a new optimization method, which includes explicit coupling terms between the orbitals and CI coefficients in the Hessian matrix. Formally, this leads to quadratic convergence of the microiterations at the expense of a higher computational cost of each microiteration. The last method is based on the alternating optimization but includes the coupling approximately using a QN approach.

III. UNCOUPLED OPTIMIZATION

In the WMK method,[19,20] the optimization of the orbitals and the CI coefficients in the microiterations is decoupled. This corresponds to an alternating optimization of the energy $E^{(2)}(\mathbf{T},\mathbf{c})$ with respect to **T** and **c**. In the following, we review the microiteration algorithm of the WMK method and describe some improvements.

In the first microiteration of each macroiteration, the CI coefficients obtained in the internal optimization step are used for computing the density matrices and optimizing the orbitals. The subsequent microiterations start with an update of the CI coefficients and density matrices, which is followed by an update of the orbitals. Both parts can involve one or more optimization steps, as will be described in the following.

A. Solving the CI eigenvalue equation

Assume that we have a set of second-order integrals as defined in Eqs. (22)–(24). For $\mathbf{T} = \mathbf{0}$, these are the initial integrals of the current macroiteration. To avoid the construction and diagonalization of the Hamiltonian matrix, the eigenvalue equation (26) is solved iteratively using a subspace optimization algorithm,[19] which can be viewed as an improved multistate Davidson method.[93] Only one or a few eigenvectors, corresponding to the states included in the state-averaging, are required.

We divide the iterative subspace into two orthogonal spaces: the primary (P) space and its complement, the Q-space. The P-space contains the CSFs with lowest energies and is selected using an energy threshold T_P as $H_{PP} - H_{00} \le T_P$, where H_{00} is the smallest diagonal element of the Hamiltonian. The assumption is that all required N_{av} states can be reasonably well described in the P-space; i.e., the contributions of the P-space CSFs dominate the wavefunctions. The Q-space is treated by a set of expansion vectors, which are linear combinations of the Q-space CSFs.

In each sub-microiteration, the eigenvalue equation is solved in a reduced space $\mathbb{C}$. The initial space $\mathbb{C}$ includes the N_p P-space CSFs and N_{q} Q-space expansion vectors $\bar{C}_{Q\alpha}$, which are obtained from the current guess of the CI-vectors by setting the coefficients of all P-space CSFs to zero and re-orthonormalizing the resulting vectors. In the first macroiteration, when no CI vectors are available, $N_{\mathrm{q}} = 0$. The reduced Hamiltonian $\mathbf{H}_C$ is built in the $N_{\mathrm{c}} = N_p + N_q$ dimensional space as described in the Appendix of Ref. 19 and then diagonalized. The eigenvectors $\bar{\mathbf{U}}$ contain the coefficients of the P-space CSFs and the expansion coefficients of the Q-space vectors. In the full space, the eigenvectors are represented by the columns of a matrix **C**

$$\bar{\mathbf{H}}_{pq} = [\bar{\mathbf{U}}^\dagger \mathbf{H}_C \bar{\mathbf{U}}]_{pq} = [\mathbf{C}^\dagger \mathbf{H} \mathbf{C}]_{pq} = E_p \delta_{pq}, \tag{28}$$

where

$$\begin{aligned} C_{Pq} &= \bar{U}_{Pq}, \\ C_{Qq} &= \sum_\alpha^{N_q} \bar{C}_{Q\alpha} \bar{U}_{\alpha q}. \end{aligned} \tag{29}$$

Here, α runs over the subspace of the Q-space vectors. For convenience, we will assume that the CI-vectors of the optimized states correspond to the first N_{av} columns of **C**, i.e., $c^n_I = C_{In}$ for $n \le N_{\mathrm{av}}$. Optionally, n additional Q-space vectors can be computed as

$$r^n_Q = \sum_\alpha^{N_q} [\mathbf{H}\bar{\mathbf{C}}]_{Q\alpha} \bar{U}_{\alpha n} + \sum_P^{N_p} H_{QP} U_{Pn} - E_n C_{Qn}, \tag{30}$$

$$\bar{C}_{Q\alpha'} = -\frac{r^n_Q}{(H_{QQ} - E_n)} \quad \text{with } \alpha' = N_q + n, \tag{31}$$

followed by Schmidt-orthonormalization on the previous Q-space vectors. For the update of the reduced Hamiltonian, the product of **H** with the new Q-space vectors is required, for which efficient algorithms are well established.[56,94,95] Each iteration is in the following denoted a "CI step."

Some care is necessary when using Slater determinants rather than CSFs as a basis: to keep the spin symmetry, all Slater determinants which belong to the same orbital configuration (which are defined solely by occupation numbers) should always be kept together in one of the two subspaces, and the energy denominators H_{QQ} used in updates should be averaged for these subsets of determinants. In addition, spin projection of the expansion vectors may be necessary to avoid numerical problems. The same

considerations apply to the usual treatment of point-group symmetry via the highest Abelian subgroup; one must then ensure that higher symmetry relations are not broken through choice of incomplete spaces, and for the case of $D_{\infty h}$ and $C_{\infty v}$, this is achieved with the help of matrix elements of the angular momentum operators.

Using the updated CI vectors, the 1-RDM and 2-RDM are computed, which are needed for the subsequent orbital optimization.

B. Solving the nonlinear equations for the orbital optimization

In order to solve the nonlinear equations (18) iteratively, an update $\mathbf{T}(\mathbf{R}+\Delta\mathbf{R})$ is defined as[19]

$$\mathbf{T}(\mathbf{R}+\Delta\mathbf{R}) := \mathbf{T}(\mathbf{R}) + \mathbf{U}(\mathbf{R})\left(\Delta\mathbf{R} + \frac{1}{2}\Delta\mathbf{R}^2 + \cdots\right). \tag{32}$$

The determination of the step $\Delta\mathbf{R}$ at the expansion point $E^{(2)}(\mathbf{T})$ requires a second-order technique that is robust enough to deal with the nonconvex problem. We employ the augmented Hessian (AH) method,[84] which determines the update by solving the eigenvalue problem

$$\begin{pmatrix} 0 & \mathbf{g}^\dagger \\ \mathbf{g} & \mathbf{h}/\lambda \end{pmatrix}\begin{pmatrix} 1/\lambda \\ \mathbf{x} \end{pmatrix} = \epsilon\begin{pmatrix} 1/\lambda \\ \mathbf{x} \end{pmatrix}, \tag{33}$$

where $\mathbf{g}$ is the gradient, $\mathbf{h}$ is the Hessian, and $\mathbf{x}$ is a vector containing all nonredundant parameters in $\Delta\mathbf{R}$. The step-length is controlled by the damping parameter λ to keep the solution $\mathbf{x}$ within a trust radius. This is equivalent to solving the NR equations with an adaptive shift ϵ of the diagonal elements of the Hessian

$$\mathbf{g} + (\mathbf{h} - \epsilon\mathbf{1})\mathbf{x} = 0 \quad \text{with} \quad \epsilon = \lambda^2\mathbf{g}^\dagger\mathbf{x}. \tag{34}$$

It can be shown that the shifted Hessian is positive definite for $\lambda \geq 1$, and therefore convergence toward an energy minimum can always be enforced by a sufficiently large value of λ.

The gradient $\mathbf{g}$ and the Hessian $\mathbf{h}$ for a fixed $\mathbf{U} = \mathbf{1} + \mathbf{T}$ are derived by inserting Eq. (32) into Eq. (9), differentiating with respect to ΔR_{rk}, and evaluating the derivatives at $\Delta\mathbf{R} = 0$. This yields

$$\tilde{g}_{rk} = \left(\tilde{A}_{rk} - \tilde{A}_{kr}\right), \tag{35}$$

$$\tilde{h}_{rk,sl} = (1-\tau_{rk})(1-\tau_{sl})\left[\tilde{G}^{kl}_{rs} - \frac{1}{2}\delta_{kl}(\tilde{A}_{rs}+\tilde{A}_{sr})\right], \tag{36}$$

with

$$\begin{aligned} \tilde{A}_{rk} &= [\mathbf{U}^\dagger\mathbf{B}]_{rk}, \qquad \tilde{A}_{ra} = 0, \\ \tilde{G}^{kl}_{rs} &= (\mathbf{U}^\dagger\mathbf{G}^{kl}\mathbf{U})_{rs}. \end{aligned} \tag{37}$$

The operator τ_{rs} permutes the indices r and s. Here, rk (r > k) and sl(s > l) are considered as composite indices of the vector $\tilde{\mathbf{g}}$ and the matrix $\tilde{\mathbf{h}}$, and the solution $\mathbf{x}$ contains the corresponding elements ΔR_{rk}. The tilde indicates that the gradient and the Hessian are evaluated for a given $\mathbf{U}$. The step-length is controlled by the damping parameter λ to keep the update within a trust radius of the second-order expansion in $\Delta\mathbf{R}$. This guarantees convergence toward a minimum. The relation between the step-length and the damping factor λ can be very steep and the interval for suitable λ can be extremely small. To avoid too short steps, we determine the damping parameter with a bisection method, which succeeds in finding a suitable λ with high precision in a few iterations.

The AH eigenvalue equation is solved iteratively, yielding an update $\Delta\mathbf{R}$. Then $\mathbf{T}$ is updated according to Eq. (32), and the process is repeated until some convergence criterion is reached (cf. Sec. III C). We denote the calculation of each update $\Delta\mathbf{R}$ an "orbital step". The iterations needed to solve the AH eigenvalue equations within each step are denoted "sub-microiterations". The process is summarized in Fig. 2.

In the original WMK method, the AH equations were solved by a simple Davidson procedure, and convergence was often slow and difficult to achieve. In our current method, we accelerate solving the AH equations by introducing a P-space, similar to the method described in Sec. III A for the CI eigenvalue problem. Here, the P-space captures the dominant orbital rotations, and the gradient $\tilde{\mathbf{g}}$ [Eq. (35)] as well as the Hessian $\tilde{\mathbf{h}}$ [Eq. (36)] are computed explicitly in this subspace. We have chosen the following criteria for selecting rotations for the P-space: first, all rotations for which the diagonal elements of the Hessian are negative are included. Since the eigenvalues of the Hessian are smaller than its smallest diagonal elements, this captures the most critical rotations. Second, the remaining rotations are sorted according to the magnitude of $|\tilde{g}_{rk}/\tilde{h}_{rk,rk}|$ and added in the order of decreasing values until a maximum P-space dimension N_p is reached. Many tests have shown that the value of N_p is not very critical, and by default, we have chosen $N_p = 200$ as a good compromise between speed and robustness of convergence and computational cost; the overhead of the P-space approach includes the explicit construction of the N_p^2 Hessian matrix elements and diagonalization of the subspace matrix (the cost scales as N_p^3). For up to several hundred P-space rotations, it is essentially negligible. The improvement through the rotational P-space will be demonstrated in Sec. VI. On the average, typically 5 sub-microiterations are needed in the AH method to reduce the gradient by a factor of 10. This threshold may not be strict enough to reach quadratic convergence in the optimization steps needed to solve Eq. (18) (for fixed CI coefficients), but since the solution also depends on the CI coefficients, higher accuracy does not lead to improved overall convergence.

The calculation of the quantities $\tilde{\mathbf{G}}^{kl}_{rs}$ defined in Eq. (37) would formally scale as $n^2_{occ}n^3_{orb}$. However, they are explicitly computed only for the P-space part of the Hessian, which involves only relatively few orbitals r, s and is therefore fast. For the iterative solution of the AH equation, one needs the products with ΔR_{sl}, which can be obtained as

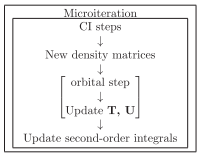

FIG. 2. Optimization scheme of a single microiteration, using the uncoupled optimization method. Several orbital steps with subsequent update of **T** (in brackets) can follow each other.

$$\sum_l [\tilde{\mathbf{G}}^{kl}\Delta\mathbf{R}]_{rl} = \left[\mathbf{U}^\dagger \sum_l \mathbf{G}^{kl}(\mathbf{U}\Delta\mathbf{R})\right]_{rl}. \tag{38}$$

This scales only as $n_{occ}^2 n_{orb}^2$ and can be well parallelized by distributing the matrices $\mathbf{G}^{kl}$ over the processors.

Finally, we note that with the WMK method sometimes different convergence behavior was observed if the program is run in parallel or on different machines. The reason for this issue is an ill-defined behavior of the AH method when the lowest eigenvalues of the (reduced) augmented Hessian matrix are degenerate. In this case, the specific eigenvector which is used to determine the update is not uniquely defined and the results can vary strongly, depending on tiny numerical changes in the augmented Hessian matrix. In the end, this can result in completely different solutions.

A remedy for this problem is to construct the update within the degenerate subspace of eigenvectors. After diagonalization of the reduced AH matrix, the degenerate eigenvectors are rotated by an orthogonal transformation such that in one (normalized) eigenvector the first entry is maximized. This is done successively by 2×2 rotations of the degenerate eigenvectors. The rotation angle α between two vectors is given by

$$\tan\alpha_1 = \frac{v_1}{v_2} \quad \text{or} \quad \tan\alpha_2 = \frac{v_2}{v_1}, \tag{39}$$

where v_1 and v_2 are the first entries of the two vectors. A rotation with either α_1 or α_2 maximizes the first entry, while the other one minimizes it. The solution with the largest first entry is kept and rotated with the next degenerate eigenvector. This is repeated until convergence is reached. By doing so, a unique update vector is obtained in the degenerate subspace, eliminating the chaotic behavior and leading to a numerically stable and reproducible solution.

C. Optimization strategy

The described orbital and CI optimizations are performed in turn. Through this alternating optimization, an indirect coupling is introduced, which leads to first-order convergence of the microiterations and second-order convergence of the macroiterations.

There are various options for fine-tuning the microiterations by choosing the number of orbital and CI steps in each microiteration. These depend on the initial orbital gradient in the current macroiteration as well as on the relative cost of the orbital and CI steps. In the initial macroiterations, when the orbital gradient is still large, it is often advantageous to carry out several orbital steps before the next CI step. In later macroiterations, the frequency between AH and CI steps is increased, and single orbital and CI steps in each microiteration are sufficient.

With large active spaces, the CI steps will be most expensive and should therefore be minimized. On the other hand, in molecules with many inactive orbitals and/or large basis sets, the orbital steps may dominate. The basic concept is to carry out in each microiteration only a single step for the more expensive problem, while more accurately converging the cheaper problem. The most suitable optimization strategy is determined automatically, depending on the total number of orbitals n_{orb}, the number of active orbitals n_{act}, the number of closed-shell orbitals n_{cl}, the number of CI states N_{av}, and the number of CSFs (Slater determinants) n_{CI}. We compare the approximate computational cost of a single CI step ($\mathcal{O}_{CI} = n_{CI}N_{av}n_{act}^4$) with the generation of the $\mathbf{G}^{ij}$ matrices, which scales approximately as $\mathcal{O}_{orb} = [n_{act}^4 + n_{act}^2 n_{cl} + n_{cl}^2/2]n_{orb}^2$. Our algorithm depends on the ratio $q = \mathcal{O}_{CI}/\mathcal{O}_{orb}$. As discussed in the Introduction, the current method is optimized for a reasonably small number of inactive orbitals. If $n_{cl} < n_{act}^2$, the contributions of n_{cl} can be neglected, and the ratio simplifies to

$$q = \frac{n_{CI}N_{av}}{n_{orb}^2}. \tag{40}$$

We found $q = 10$ to be a reasonable threshold for switching from an orbital dominating ($q \le 10$) to a CI dominating ($q > 10$) algorithm. The choice may not be optimal near the switching point, but typically the computation times of both methods are close at $q = 10$. More details about the convergence parameters can be found in the supplementary material.

IV. COUPLED OPTIMIZATION OF THE SECOND-ORDER ENERGY APPROXIMATION

The uncoupled optimization of the orbitals and CI coefficients outlined in Sec. III yields only first-order convergence of the microiterations. While this works well in most cases, sometimes situations are encountered where the decay of the energy and gradient is very slow and the microiterations are difficult to converge.

In this section, we address this problem by including explicitly the coupling of the orbital rotations and changes in the CI vectors in the AH equations, leading to second-order convergence of the microiterations. In order to minimize the computational cost, only the coupling of the (full) orbital space with the reduced CI space $\mathbb{C}$ introduced in Sec. III is included. Initially, the subspace $\mathbb{C}$ contains the P-space and at least one Q-space vector. By iteratively increasing the number of Q-space vectors in $\mathbb{C}$, the coupling with the full CI space can be approached.

The CI coefficients are optimized within the reduced space $\mathbb{C}$ by an orthogonal transformation $\mathbf{V}$ of the orthonormal basis of expansion vectors $\mathbf{C}$. The initial vectors in $\mathbf{C}$ diagonalize the Hamiltonian in this subspace, but this is not the case any more when further vectors are added in the sub-microiterations (see below). Similar to Eq. (3), this transformation is parameterized by an antisymmetric matrix $\mathbf{S}$

$$\mathbf{V} = \exp(\mathbf{S}), \qquad \mathbf{S} = -\mathbf{S}^\dagger \tag{41}$$

so that a first-order update of the CI-coefficients is

$$\Delta\mathbf{C} = \mathbf{CS}. \tag{42}$$

Only rotations with the N_{av} optimized states are required, and all other rotations are redundant and set to zero, i.e., $S_{pq} = 0$ for $q > N_{\text{av}}$. As we will see below, rotations between states with equal weights are also redundant and are excluded. If all nonzero weights are identical, only rotations $S_{pm} = -S_{mp}$ for $p > N_{\text{av}}$, $m \le N_{\text{av}}$ remain. This parameterization automatically takes care of the orthonormalization conditions of the CI-vectors.

We now minimize the state-averaged energy expression

$$E_{\text{T}}(\mathbf{S}, \Delta\mathbf{R}) = \sum_{n=1}^{N_{\text{av}}} W_n[\exp(-\mathbf{S})\bar{\mathbf{H}}(\Delta\mathbf{R})\exp(\mathbf{S})]_{nn} \tag{43}$$

with respect to $\Delta\mathbf{R}$ and $\mathbf{S}$. Here, the subscript $\mathbf{T}$ indicates that the second-order energy expression $E^{(2)}(\mathbf{T})$ with the update formula (32) is employed, and the expansion point is defined by a fixed $\mathbf{T}$. Expanding Eq. (43) up to second order in S_{pm} and ΔR_{ri} yields formally

$$E^{(2)}_{\mathbf{T}}(\mathbf{S},\Delta\mathbf{R}) = E^{(2)}(\mathbf{T}) + \tilde{\mathbf{g}}_c^\dagger \mathbf{x}_c + \tilde{\mathbf{g}}_o^\dagger \mathbf{x}_o + \frac{1}{2}\mathbf{x}_c^\dagger \tilde{\mathbf{h}}_{cc}\mathbf{x}_c + \frac{1}{2}\mathbf{x}_o^\dagger \tilde{\mathbf{h}}_{oo}\mathbf{x}_o + \mathbf{x}_o^\dagger \tilde{\mathbf{h}}_{oc}\mathbf{x}_c, \tag{44}$$

where $E^{(2)}(\mathbf{T})$ is the second-order energy at the expansion point. $\mathbf{x}_o$ and $\mathbf{x}_c$ are vectors which contain the nonredundant parameters ΔR_{ri} $(r > i)$ and S_{pm} $(p > m)$, respectively. $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{h}}$ are the corresponding first and second derivatives for $\mathbf{S} = 0$, $\Delta\mathbf{R} = 0$, and the orbital and CI sub-blocks are indicated by the subscripts o and c, respectively. The tildes over the symbols indicate that they are evaluated with a fixed $\mathbf{T}$. Minimizing with respect to $\mathbf{x}$ and subsequently replacing the Hessian by $\tilde{\mathbf{h}} - \epsilon\mathbf{1}$ with $\epsilon = \tilde{\mathbf{g}}^\dagger\mathbf{x}$ yields the AH equations

$$\begin{pmatrix} -\epsilon & \tilde{\mathbf{g}}_c^\dagger & \tilde{\mathbf{g}}_o^\dagger \\ \tilde{\mathbf{g}}_c & \tilde{\mathbf{h}}_{cc} - \epsilon & \tilde{\mathbf{h}}_{co} \\ \tilde{\mathbf{g}}_o & \tilde{\mathbf{h}}_{oc} & \tilde{\mathbf{h}}_{oo} - \epsilon \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{x}_c \\ \mathbf{x}_o \end{pmatrix} = \mathbf{0}. \tag{45}$$

For simplicity, we have omitted the damping parameter λ. The orbital parts of the gradient and Hessian are given by Eqs. (35) and (36), respectively. For the CI part, one obtains $(p > m, q > n)$,

$$[\tilde{\mathbf{g}}_c]_{pm} = 2\big[W_m - W_p\big]\tilde{H}_{pm}, \tag{46}$$

$$[\tilde{\mathbf{h}}_{cc}]_{pm,qn} = (1-\tau_{pm})(1-\tau_{qn})\Big[\tilde{H}_{pq}\delta_{mn}\big(2W_m - W_p - W_q\big)\Big], \tag{47}$$

where we assume that $W_p = 0$ for $p > N_{\mathrm{av}}$ and $\tilde{\mathbf{H}} \equiv \mathbf{C}^\dagger\mathbf{H}_\mathbf{T}\mathbf{C}$, with $\mathbf{H}_\mathbf{T} \equiv \mathbf{H}^{(2)}(\mathbf{T})$ as given in Eq. (21). The gradient includes the off-diagonal entries of the reduced Hamiltonian, which become zero if $\tilde{\mathbf{H}}$ is diagonal. The first derivatives for rotations between the N_{av} states with the same weights are zero despite nonzero off-diagonal values. Similarly, the corresponding Hessian elements are zero, which reflects the fact that these rotations are redundant. If we assume that the weights for all states m, n are the same, and that $\tilde{H}_{mn} = E^{(2)}_m\delta_{mn}$, the CI gradient and Hessian simplify to

$$\big[\tilde{\mathbf{g}}_c\big]_{pm} = 2W_m\tilde{H}_{pm}, \tag{48}$$

$$\big[\tilde{\mathbf{h}}_{cc}\big]_{pm,qn} = 2W_m\delta_{nm}(\tilde{H}_{pq} - E^{(2)}_m\delta_{pq}) \tag{49}$$

for $W_p = W_q = 0$. Thus, there are no couplings of rotations for different states, and $\tilde{\mathbf{h}}_{cc}$ has a block-diagonal structure, with one block for each optimized state. Since the gradient includes only the off-diagonal elements between the N_{av} states n and the other reduced basis vectors p, only the subspace spanned by the N_{av} states is optimized. Hence, the off-diagonal elements between the N_{av} states are not necessarily vanishing after the optimization. This is fixed by diagonalizing the Hamiltonian in the basis of the N_{av} states after updating the CI vectors with $\mathbf{S}$.

For the sake of simplicity, we have in the following not separated the inactive and active orbital spaces, and in the remainder of this section, the indices i, j, k, l are assumed to run over all occupied orbitals. Explicit formulae which take into account the simplifications for the closed-shell orbitals are given in the Appendix. The coupling part between the CI and orbital rotations is

$$[\tilde{\mathbf{h}}_{oc}]_{ri,pm} = 4(W_m - W_p)\Big[\tilde{A}^{pm}_{ri} - \tilde{A}^{pm}_{ir}\Big]. \tag{50}$$

Here, $\tilde{\mathbf{A}}^{pm}$ is defined as

$$\tilde{A}^{pm}_{ri} = \sum_j^{\mathrm{occ}}(\mathbf{U}^\dagger\mathbf{h}\mathbf{U})_{rj}D^{pm}_{ij} + \sum_{jkl}^{\mathrm{occ}}(\mathbf{U}^\dagger\mathbf{J}^{kl}\mathbf{U})_{rj}D^{pm}_{ij,kl} + 2\sum_{jkl}^{\mathrm{occ}}(\mathbf{U}^\dagger\mathbf{K}^{kl}\mathbf{T})_{rj}D^{pm}_{ik,jl}, \tag{51}$$

where D^{pm}_{ij} and $D^{pm}_{ij,kl}$ are transition density matrices between the N_{av} states and the other reduced subspace vectors of $\mathbf{C}$, e.g.,

$$D^{pm}_{ij} = \sum_{IJ} C_{Ip}D^{IJ}_{ij}C_{Jm} = \sum_I C_{Ip}D^{Im}_{ij}. \tag{52}$$

The coupled AH equations are solved using the P-space variant of the Davidson method. For simplicity, we will first consider the case that the orbital parts of the Hessian are explicitly computed. The diagonalization of the AH matrix (45) then yields the rotation parameters ΔR_{ri} and S_{pm}. In order to approach the coupling with the full CI space, the reduced space has to be augmented with further Q-space vectors. For this, we need to compute the CI residuals directly in the full CSF basis

$$r^n_Q = [\mathbf{HC}(\mathbf{1}+\mathbf{S})]_{Qn} - E_n[\mathbf{C}(\mathbf{1}+\mathbf{S})]_{Qn} - \epsilon[\mathbf{CS}]_{Qn} + 2\sum_{ri}\tilde{A}^{Qn}_{ri}\Delta R_{ri} - 2\mathbf{C}_{Qn}\sum_{ri}\tilde{A}^{nn}_{ri}\Delta R_{ri}, \tag{53}$$

where ϵ is the eigenvalue obtained from the reduced AH problem. The coupling terms in Eq. (53) are evaluated as

$$\sum_{r,i}\tilde{A}^{Qn}_{ri}\Delta R_{ri} = [\widetilde{\mathbf{H}}^{(2)}\mathbf{c}^n]_Q \quad \text{and} \quad \sum_{ri}\tilde{A}^{nn}_{ri}\Delta R_{ri} = \tilde{E}^{(2)}_n, \tag{54}$$

where $\widetilde{\mathbf{H}}^{(2)}$ and $\tilde{E}_n$ are computed with transformed integrals

$$\tilde{h}_{ij} = \frac{1}{2}(1+\tau_{ij})(\Delta\mathbf{R}^\dagger\mathbf{U}^\dagger\mathbf{h}\mathbf{U})_{ij}, \tag{55}$$

$$\widetilde{(ij|kl)} = \frac{1}{2}(1+\tau_{ij})\big[(\Delta\mathbf{R}^\dagger\mathbf{U}^\dagger\mathbf{J}^{kl}\mathbf{T})_{ij} + (\Delta\mathbf{R}^\dagger\mathbf{U}^\dagger\mathbf{K}^{jk}\mathbf{U})_{il} + (\Delta\mathbf{R}^\dagger\mathbf{U}^\dagger\mathbf{K}^{jl}\mathbf{U})_{ik}\big]. \tag{56}$$

Using the residuals in Eq. (53), N_{av} new Q-space expansion vectors $\tilde{C}_{Q\alpha}$ can be obtained by preconditioning, as shown in Eq. (31). The new Q-space vectors are orthogonalized on the previous vectors in $\mathbf{C}$ and then appended to the coefficient matrix $\mathbf{C}$. The new rows in the Hessian $\tilde{\mathbf{h}}_{cc}$ can be computed using standard direct-CI routines. The new coupling matrix elements $[\tilde{\mathbf{h}}_{oc}]_{ri,pq}$ are computed with Eq. (51). Subsequently, the AH equations are solved again (sub-microiterations). This is repeated until the second-order expansion in Eq. (44) is converged, and the AH residual is small enough.

In practice, the orbital part in the AH equations is also split into a P-space and a Q-space, and only the P-space rotations are explicitly included in the reduced space. For the Q-space rotations, one can compute in each iteration a Q-space update vector ΔR^q_{ri}, as

described in Sec. III. These Q-space vectors are linearly combined using the coefficients $(\mathbf{x}_o)_q$ obtained by solving the AH equations in the reduced space.

After convergence of the AH calculation, the CI vectors are updated by transforming $\mathbf{C}$ with $\mathbf{V} = \exp(\mathbf{S})$ and $\mathbf{T}$ is updated according to Eq. (32). The integrals are transformed according to Eqs. (22)–(24), and the next microiteration is started by a pure CI optimization with the new second-order Hamiltonian $\mathbf{H}^{(2)}(\mathbf{T})$ and the transformed CI-vectors to create a new reduced space $\mathbb{C}$. Afterwards, the coupled AH equations are solved again as described above. This procedure shows quadratic convergence, if sufficient Q-space vectors $C_{Q\alpha}$ are added.

A disadvantage of the described coupled optimization method is that for each state n and Q-space expansion vector $[\mathbf{c}^\alpha]_Q \equiv C_{Q\alpha}$, the transition density matrices $D^{n\alpha}_{ij}$ and $D^{n\alpha}_{ij,kl}$ are required. Furthermore, for each additional expansion vector obtained in the sub-microiterations, the products $\tilde{\mathbf{H}}\mathbf{c}^n$ and $\mathbf{H}_\mathbf{T}\mathbf{c}^\alpha$ have to be evaluated. It is possible to reduce the number of transition density matrices in the multistate case by excluding rotations $S_{n\alpha}$ when the corresponding matrix elements $(\mathbf{c}^n\mathbf{H}_\mathbf{T}\mathbf{c}^\alpha)$ are tiny. This avoids the computation of the transition densities between $\mathbf{c}^n$ and $\mathbf{c}^\alpha$. We found in many tests that the effect of these approximations is very small. It is therefore used by default but can be disabled by a program option. In the uncoupled method, only N_{av} vectors $\mathbf{H}\mathbf{c}^m$ per Davidson iteration and one set of state-averaged density matrices are required per CI update. As shown in Sec. VI, the coupled method usually reduces the number of orbital updates but requires a larger number of $\mathbf{H}\mathbf{c}$ and density evaluations than the uncoupled method.

In Sec. VI, we distinguish between the coupled method without and with additional expansion vectors. The former is denoted "Partially Coupled" and the one with additional expectation vectors as "Fully Coupled."

V. QUASI-NEWTON COUPLING

As will be shown in Sec. VI, inclusion of the orbital-CI coupling in the microiterations can lead to a significant improvement of the convergence. This even applies to the macroiterations, in cases where the coupling is strong and convergence of the microiterations is difficult to achieve with the uncoupled solver. However, the price one has to pay for inclusion of the coupling is a higher cost for the CI work since many additional transition density matrices have to be computed. In this section, we simplify the coupled treatment by a quasi-Newton approach.[28,32]

We assume that we are already reasonably close to convergence and that the parameters $\mathbf{x}_c$ and $\mathbf{x}_o$ in Eq. (45) can be obtained by solving the linear equations

$$\begin{pmatrix} \tilde{\mathbf{h}}_{cc} & \tilde{\mathbf{h}}_{co} \\ \tilde{\mathbf{h}}_{oc} & \tilde{\mathbf{h}}_{oo} \end{pmatrix} \begin{pmatrix} \mathbf{x}_c \\ \mathbf{x}_o \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{g}}_c \\ \tilde{\mathbf{g}}_o \end{pmatrix} = \mathbf{0}. \tag{57}$$

This can formally be rewritten as

$$\tilde{\mathbf{g}}_o - \tilde{\mathbf{h}}_{oc}\tilde{\mathbf{h}}_{cc}^{-1}\tilde{\mathbf{g}}_c + [\tilde{\mathbf{h}}_{oo} - \tilde{\mathbf{h}}_{oc}\tilde{\mathbf{h}}_{cc}^{-1}\tilde{\mathbf{h}}_{co}]\mathbf{x}_o = 0. \tag{58}$$

If the initial CI-gradient is zero, $\tilde{\mathbf{g}}_c = \mathbf{0}$, this differs from the pure orbital equation only by replacing the orbital Hessian $\mathbf{h}_{oo}$ by an effective Hessian $\tilde{\mathbf{h}}_{oo} = \tilde{\mathbf{h}}_{oo} - \tilde{\mathbf{h}}_{oc}\tilde{\mathbf{h}}_{cc}^{-1}\tilde{\mathbf{h}}_{co}$. Of course, in cases with large CI spaces, the coupling part cannot be computed explicitly, but the idea is to use $\tilde{\mathbf{h}}_{oo}$ as a starting guess and to update it using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) update formula[96,97]

$$\bar{\mathbf{h}}^k = \bar{\mathbf{h}}^{k-1} + \frac{\mathbf{y}^k\mathbf{y}^{k\dagger}}{\mathbf{y}^{k\dagger}\mathbf{s}^k} - \frac{\bar{\mathbf{h}}^{k-1}\mathbf{s}^k\mathbf{s}^{k\dagger}\bar{\mathbf{h}}^{k-1\dagger}}{\mathbf{s}^{k\dagger}\bar{\mathbf{h}}^k\mathbf{s}^k} \tag{59}$$

with

$$\mathbf{y}^k = \tilde{\mathbf{g}}_o^k - \tilde{\mathbf{g}}_o^{k-1} \quad \text{and} \quad \mathbf{s}^k = \mathbf{x}_o^{k-1}. \tag{60}$$

Instead of solving the linear equation, one can equally well employ this updated Hessian in the AH approach. The index k denotes subsequent orbital steps, and between each of these the CI coefficients are updated.

The algorithm can be summarized as follows: Solving the AH equations yields $\mathbf{x}_o^k \equiv \Delta\mathbf{R}$. This is then used to update $\mathbf{T}$, and the second-order integrals are recomputed and used in the next CI optimization. Here, we ensure that the norm of the CI residual becomes substantially lower than the last orbital gradient to justify the approximation $\mathbf{g}_c = 0$ in Eq. (58). Afterwards, the RDMs and a new orbital gradient are calculated, followed by the next BFGS update of the Hessian. This is repeated until convergence to a certain threshold is reached. Note that the initial orbital Hessian $\tilde{\mathbf{h}}_{oo}$ is normally not recomputed in each microiteration; its change due to $\Delta\mathbf{R}$ is assumed to be accounted for by the BFGS update. Only if the averaged density strongly changes or the energy increases in an iteration, the orbital Hessian is recalculated and the BFGS method is restarted. Furthermore, to avoid numerical instabilities, the BFGS update is omitted if one of the denominators becomes tiny.

The construction of the full approximate Hessian is not needed since the iterative diagonalization of the AH matrix requires only products of the Hessian with the Q-space vectors $\mathbf{q}$. It is only necessary to store in each iteration k the vectors

$$\mathbf{v}_1^k = \mathbf{y}^k \quad \text{and} \quad \mathbf{v}_2^k = \tilde{\mathbf{h}}^{k-1}\mathbf{s}^k \tag{61}$$

and the denominators

$$d_1^k = \mathbf{v}_1^{k\dagger}\mathbf{s}^k \quad \text{and} \quad d_2^k = \mathbf{v}_2^{k\dagger}\mathbf{s}^k. \tag{62}$$

Finally, the product reads

$$\tilde{\mathbf{h}}^k\mathbf{q} = \tilde{\mathbf{h}}^0\mathbf{q} + \sum_{i=1}^{k}\frac{\mathbf{v}_1^i(\mathbf{v}_1^{i\dagger}\mathbf{q})}{d_1^i} - \sum_{i=1}^{k}\frac{\mathbf{v}_2^i(\mathbf{v}_2^{i\dagger}\mathbf{q})}{d_2^i}, \tag{63}$$

where $\tilde{\mathbf{h}}^0\mathbf{q}$ is the product of the initial orbital Hessian with a Q-space vector $\mathbf{q}$. This Hessian-vector product is also required for the computation of $\mathbf{v}_2^i$. The additional cost for storing the BFGS update vectors and the additional computations in the Hessian vector product is negligible.

The whole process differs from the uncoupled orbital optimization (with focus on minimizing the orbital steps) only by the BFGS update of the Hessian instead of recomputing it after each orbital step. Therefore, the computational effort is even slightly lower than for the uncoupled optimization. However, due to the approximations in the BFGS update, it is not guaranteed that the overall convergence is improved by the QN approach. Nevertheless, we found that in cases where the microiterations are slowly convergent, the

QN coupling accelerates the convergence of the microiterations considerably. This will be demonstrated in Sec. VI.

VI. CALCULATIONS

The MCSCF solvers described in Secs. III–V have been implemented in the MOLPRO package.[86] All of the different optimization strategies presented in this paper can be chosen using options as described in the user's manual. In this section, we compare the convergence properties of these methods for excitation energies of 21 aromatic molecules[91] and five metal complexes. All benchmark systems are based on already published systems, and we only study the convergence behavior of the presented methods. We compare the following properties of the methods: (i) the number of macroiterations, (ii) the number of microiterations, (iii) the number of orbital optimization steps (updates with $\Delta\mathbf{R}$), (iv) the total number of sub-microiterations for the AH matrix diagonalization, and (v) the CI work. The latter is taken to be the number of all $\mathbf{H}c$, density, and transition density matrix evaluations, which are similarly expensive.

In the tables, we compare these numbers for the various optimization algorithms. "Uncoupled (CI)" and "Uncoupled (Orb)" denote the uncoupled algorithms, which focus on minimizing the number of CI and orbital steps, respectively. "Partially Coupled" includes the orbital-CI coupling only via the CI P-space and the initial Q-space vector(s), while "Fully Coupled" adds further Q-space vectors in each sub-microiteration (cf. Sec. IV). WMK denotes the earlier implementation in Molpro,[19,20] which follows a similar optimization strategy as "Uncoupled (CI)."

The Molpro default convergence criteria were used: in all cases, the final energy is converged to better than 10^{-8} E_h, and the norm of the orbital gradient is smaller than 10^{-4}.

A. Simple aromatic systems

Our first calculations are based on a benchmark set introduced by Menezes *et al.*[91] to demonstrate the accuracy of their PNO-CASPT2 method. It contains the calculation of the first π-π^* excitation energies of 21 aromatic systems using state-averaged CASSCF. The geometries are obtained from the supplementary material of Ref. 91. The active spaces include the π orbitals of the aromatic systems and capture various sizes from CAS(6,6) up to CAS(12,12). For niacinamide, niacin, and picolinic acid, we removed one nearly doubly occupied orbital from the active space and changed the space from CAS(10,9) to CAS(8,8). All calculations used density fitting[98,99] for the two-electron integral transformation. The starting orbitals were obtained from a density-fitted Hartree Fock calculation, and the aug-cc-pVDZ and aug-cc-pVTZ basis sets[100] were used. All calculations were done without using symmetry.

Table I shows the numbers of macroiterations, microiterations, orbital optimization steps, sub-microiterations, and the CI work. All numbers are summed over the 21 calculations. By comparison of the WMK and "Uncoupled (CI)" numbers, one can see that the introduction of the rotational P-space significantly reduces the number of orbital and CI steps, as well as the number of macroiterations.

The orbital focused "Uncoupled (Orb)" algorithm further reduces the number of orbital steps, without strongly increasing the CI work. In this case, the number of sub-microiterations is reduced by a factor of 3–4 relative to the WMK method. In general, we observed that better and more stable convergence of the CI steps significantly stabilizes overall convergence and often also reduces the number of orbital steps.

For these relatively simple systems, no significant reductions of the number of macroiterations or microiterations are achieved by the coupled methods. The slight reduction of the number of orbital steps and sub-microiterations is outweighed by a strong increase of the CI work, which is due to the many transition density evaluations. As expected, the convergence and computational effort of the QN method is for these molecules very similar to the "Uncoupled (Orb)" method.

B. Fe(II)-porphine

The next example is the Fe(II)-porphine transition metal complex.[44,75,101–116] Porphine is the simplest representative of the porphyrin group, and its derivatives are important in the human metabolism. The focus of most computational studies is the correct ordering of the spin states.[103,107–111] Recently, CASSCF computations with approximative CI methods have been published,[44,75,115,116] investigating the ordering for various active spaces. The authors concluded that very large active spaces are needed for a proper description of the excitation energies.[116] Nevertheless, we use a

TABLE I. Results for the aromatic benchmarks set[91] summed over all 21 calculations. The number of macroiterations, microiterations, orbital steps, sub-microiterations of the AH diagonalization, and a measure for the CI work are presented. The latter is the sum of **Hc**, density matrix, and transition density matrix (1-RDMs and 2-RDMs) evaluations.

	aug-cc-pVDZ					aug-cc-pVTZ				
Solver	Macro	Micro	Orbital steps	Sub-micro	CI work	Macro	Micro	Orbital steps	Sub-micro	CI work
WMK	75	621	791	5447	5 388	75	616	788	5475	5 376
Uncoupled (CI)	70	391	629	2271	4 290	70	395	634	2438	4 312
Uncoupled (Orb)	69	376	376	1477	4 533	69	378	378	1573	4 542
Partially Coupled	69	356	356	1537	9 853	69	360	360	1665	9 955
Fully Coupled	69	348	348	1501	15 436	69	352	352	1607	15 903
QN-Coupled	69	383	383	1561	4 585	69	388	388	1670	4 604

small active space for studying the CASSCF convergence for this system.

We reproduced the calculations of various singlet, triplet, and quintet states of this system presented in a recent publication by Sun *et al.*,[44] where slow convergence of the uncoupled method was reported. All calculations were performed with the CAS(8,11) active space including the 3d and 4d orbitals as well as one orbital for the nitrogen atoms. The latter forms bonding and antibonding MOs with the Fe($3d_{x^2-y^2}$) orbital.[111] We used the triplet geometry[110] and the cc-pVDZ (vdz), cc-pVTZ (vtz), and cc-pVQZ (vqz) basis sets,[117,118] as in Ref. 44. In the calculations, the D_{2h} point-group symmetry was exploited. Starting orbitals were produced using the "Automated Construction of Molecular Active Spaces from Atomic Valence Orbitals" (AVAS)[70] approach, based on ROHF orbitals for the $^3B_{1g}$ state. The target atomic valence orbitals were Fe(3d) and Fe(4d) approximated in the cc-pVTZ basis, and the AVAS truncation threshold was 0.3. This automatically produced the active space of 11 orbitals.

Using these starting orbitals, we carried out a state-averaged calculation including the ^1A_g, $^3B_{1g}$, $^3B_{3g}$, ^5A_g, and $^5B_{2g}$ states. The resulting orbitals were used as starting guesses in the single-state calculations for the 5 states. The resulting energies were identical to those reported in Ref. 44, except for the ^1A_g state energy at the vqz level, where our result is 12.4 μE_h lower. In these calculations, the computational effort was clearly dominated by the two-electron integral transformation in each macroiterations for all three basis sets. Hence, the computation time is basically linearly dependent on the number of macroiterations. Calculations with the cc-pVQZ basis included the optimization of 1784 molecular orbitals, and each macroiteration took around half an hour elapsed time on five CPU cores. In these calculations, the 2-electron AO integrals were stored on disk, which is possible due to the high symmetry of the molecule. For comparison, DF calculations were also carried out, but these were much more expensive since our program does not support symmetry in the DF integral evaluation and transformation steps.

The number of macroiterations and microiterations of the state-averaged and single state calculations is shown in Table II. For this small active space, the number of CI steps is rather irrelevant, and therefore we do not show the results for the uncoupled (CI) variant [which are rather similar to the "Uncoupled (Orb)" ones anyway]. Note that in the "Uncoupled (Orb)" method the number of microiterations equals the number of orbital steps.

For the singlet and triplet states, the WMK method shows severe convergence problems, which become worse with increasing basis set. The convergence of the ^1A_g state at the vqz level is extremely slow, and the WMK calculation was aborted after 40 macroiterations without convergence. This problem is due to the swapping of a weakly occupied active orbital with a virtual orbital during the optimization. This leads to a strong coupling of the orbital and CI optimization problems.

The new "Uncoupled (Orb)" method benefits from the introduction of the rotational P-space, which results in more accurate updates of the orbital rotations and thus improves convergence of both the microiterations and macroiterations. Switching on the explicit orbital-CI coupling reduces the number of orbital steps further. However, additional expansion vectors obtained with the fully coupled scheme are required for reducing the number of macroiterations. The fully coupled method yields the lowest number of macroiterations but introduces an increase of the CI work due to the additional transition densities and expansion vectors. However, for the small active space of CAS(8,11), this has no significant impact on the computation time, and overall, the fully coupled calculations are the fastest of all variants. The improvements by the coupling are particularly evident for the vqz calculations of the singlet and triplet states. The quintet states converge smoothly with all methods.

In the last column, the results of the QN-coupling method are shown. The number of macroiterations is in most cases equal to the fully coupled optimization. The number of microiterations is somewhat larger than for the fully coupled method, but the number of microiterations and CI steps is still lower than for the "Uncoupled (Orb)" method.

C. $Cu_2O_2^{2+}$ isomerization

We now consider the isomerization of a $[(NH_3)_3Cu]_2O_2^{2+}$ complex[119] from the bis(μ-oxo) to the μ-μ^2:μ^2 peroxo structure. The computation of this system is challenging through the balance between the dynamical correlation at the bis(μ-oxo) structure and the static correlation through the biradical character at the peroxo structure. The system has been heavily investigated on the theoretical level since studies showed a strong disagreement between CASPT2 and DFT calculations.[120,121] An extensive benchmark study of Cramer *et al.*[122] compared results of various correlation methods. It was concluded that there is a good agreement between completely renormalized coupled cluster (CR-CC), MRCI, and nonhybrid DFT functionals like BLYP, while CASPT2 shows quantitatively wrong results. Finally, RASPT2 calculations[123] solved the disagreement of the perturbation theory by including the Cu 3d-4d correlation in the reference functions. In the meantime, DMRG calculations[124–126] and DMRG-CASPT2[112] results are also available.

We carried out the CASSCF calculations as described in Ref. 122 using the all-electron aug-cc-pVTZ basis set.[100] The calculation of the ^1A_g ground state in C_{2h} symmetry includes an active space of 16 electrons in 14 orbitals. The structures and more details on the active space can be found in the supplementary material of Ref. 122. The benchmark comprises six CASSCF calculations along an isomerization path from the bis(μ-oxo) (F = 0) to the peroxo (F = 100) structure. We started the calculations at the bis(μ-oxo) structure with AVAS orbitals obtained from a closed-shell Hartree Fock calculation. The AVAS calculation used the 2s, 3s, 2p, 3p orbitals of the oxygen atoms and the $3d_{yz}$ orbitals of the Cu atoms as target space (O_2 lies on the z-axis, and the Cu atoms are displaced in the y-direction; C_{2h} symmetry). The oxygen atomic target orbitals were generated using a SA-CASSCF/cc-pVTZ calculation for the three degenerate ^3P states with 2s, 3s, 2p, 3p active. The AVAS threshold was 0.3. This generated an active space of 16 electrons in 18 orbitals ($4a_g$, $5a_u$, $5b_u$, $5b_g$ in C_{2h} symmetry. Using these starting orbitals, the CASSCF(16,14) converged in 3 macroiterations. The subsequent calculations along the reaction path used the orbitals from the previous structure as the starting guess. However, despite these apparently good starting orbitals, the convergence slowed down along the reaction path, and at some points, the WMK method almost failed to converge. We found that this is due to qualitative changes in the

TABLE II. Result of the optimization of the ^1A_g, $^3B_{1g}$ and $^3B_{3g}$, ^5A_g and $^5B_{2g}$ states of the Fe(II)-porphine complex. The number of macroiterations (Mac), the number of orbital optimization steps (Orb), and the measure for the CI cost (CI) are presented. The calculations are performed with an active space of CAS(8,11). The single state calculations use the orbitals from the state-averaged calculation as a starting guess. Italic numbers indicate that no convergence could be achieved after reaching the maximum of 40 iterations.

		WMK			Uncoupled (Orb)			Partially Coupled			Fully Coupled			QN-Coupled		
State	Basis	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI
SA	vdz	8	98	1381	4	26	751	4	23	1225	4	21	1921	4	26	736
	vtz	11	132	1819	4	27	810	4	22	1163	4	22	2188	4	28	790
	vqz	12	138	1950	4	28	842	4	22	1178	4	22	2290	4	28	774
^1A_g	vdz	11	107	372	4	35	191	4	24	287	3	14	294	3	21	159
	vtz	18	171	571	4	51	247	5	35	375	3	17	380	4	30	208
	vqz	*40*	*456*	*1265*	12	379	1462	10	167	1421	6	42	965	7	121	799
$^3B_{1g}$	vdz	15	235	650	5	72	354	5	45	453	3	20	409	4	50	382
	vtz	7	105	304	4	47	261	4	27	307	3	21	436	3	40	260
	vqz	7	104	303	5	50	275	4	27	302	3	21	433	3	40	259
$^3B_{3g}$	vdz	4	39	135	4	30	149	3	20	206	3	14	253	3	21	126
	vtz	7	66	219	4	37	170	4	22	235	3	14	281	3	19	124
	vqz	8	76	245	4	40	182	4	25	256	3	15	310	3	22	131
^5A_g	vdz	3	27	85	3	16	88	3	15	138	3	15	227	3	17	91
	vtz	3	31	95	3	20	101	3	18	161	3	18	266	3	17	91
	vqz	3	32	95	3	21	104	3	18	161	3	18	274	3	20	99
$^5B_{2g}$	vdz	3	31	96	3	17	90	3	15	147	3	14	233	3	17	94
	vtz	3	31	97	3	21	104	3	18	162	3	17	278	3	22	110
	vqz	3	32	97	3	21	104	3	17	156	3	17	286	3	22	112

weakly occupied $23a_g$ and $13b_g$ orbitals, which are localized at the O–O bond. The $23a_g$ orbital changes its character from an $O(3p_\sigma)$ bonding orbital at F = 0 to an O(3s) bonding orbital at F = 100, while the $13b_g$ orbital changed from an $O(3p_\pi)$ antibonding out-of plane orbital at F = 0 to the corresponding in-plane orbital at F = 100. The strongest changes in the orbitals occur in the range F = 40–80, and this results in the convergence difficulties. This problem can be avoided by adding one additional a_g and one additional b_g orbital to the active space, resulting in a CAS(16,16) space. As can be seen in Table III, the calculations with this extended active space converge smoothly with all optimization methods. In fact, this is consistent with our long-term experience that the slow convergence of the WMK method often indicates an inappropriate active space.

Table III shows the number of macroiterations (Mac) and orbital optimization steps (Orb) as well as the CI work (CI) along the isomerization path. The resulting orbitals and energies are equal for all methods. The changes in the orbitals at F = 40–80 lead to slow convergence for all uncoupled methods. The WMK method shows extremely slow convergence, and the convergence is hardly reached at F = 60 after 40 macroiterations, where the calculation was aborted. Coupling without additional expansion vectors accelerates the convergence, but for F = 60, it is still slow. Adding further expansion vectors increases the coupling and yields the fastest convergence. In the F = 40–80 region, the additional CI work from the expansion vectors is overcompensated by the much faster convergence. The QN-coupling reaches a nearly as fast convergence in the macroiterations, while no additional CI work is required for the coupling. Hence, this solver requires the lowest total CI work, especially in the F = 40–80 region. Since this active space includes over 2×10^6 determinants, the CI optimization dominates the computation time and the QN coupling provides by far the fastest computation of the whole isomerization calculation.

The CASSCF(16,16) calculations include about 40×10^6 Slater determinants and are therefore even more strongly dominated by the CI work. In this case, the "Uncoupled (CI)" method is fastest in the initial calculation for F = 0, but overall the number of CI steps in the "Uncoupled (Orb)" and QN methods is nearly the same.

In Table IV, we present some timings for the copper complex with CAS(8,10), CAS(16,14), and CAS(16,16) active spaces. The starting orbitals for the CAS(8,10) active space were obtained using $O(2p_x,3p_x,2p_z,3p_z)$ and $Cu(3d_{yz})$ target orbitals. In the calculation with the small CAS(8,10) active space, the computational effort for the integral transformation clearly dominates, while in the large CAS(16,16) calculation, the CI work strongly dominates. The fully coupled calculation with the CAS(8,10) active space takes a few seconds less than the one with QN coupling and is therefore not shown. For the CAS(16,14) active space, the fully coupled algorithm takes about 15% more time than the QN one, while for the CAS(16,16) case it is overall by a factor of ≈1.8 times slower (not

TABLE III. Macroiterations (Mac), orbital steps (Orb), and CI work (CI) of the isomerization of $[(NH_3)_3Cu]_2O_2^{2+}$ from the bis(μ-oxo) to the peroxo structure[119] with an active space of CAS(16,14) and CAS(16,16). Italic numbers indicate that no convergence could be achieved after reaching the maximum of 40 iterations.

	WMK			Uncoupled (CI)			Uncoupled (Orb)			Partially Coupled			Fully Coupled			QN-Coupled		
F	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI
								CAS(16,14)										
0	4	55	211	4	56	198	3	25	189	4	23	328	3	19	424	3	24	195
20	4	47	211	4	58	245	4	33	247	3	19	281	3	14	333	4	30	282
40	7	86	328	4	63	233	4	46	277	4	24	324	3	15	337	3	28	238
60	36	340	1308	8	262	827	7	204	911	5	84	787	4	22	538	4	38	279
80	*40*	*542*	*1626*	13	463	1468	11	378	1624	7	161	1396	4	23	509	5	51	343
100	*40*	*462*	*1441*	8	259	835	7	214	928	5	74	664	3	18	413	4	35	223
SUM	131	1532	5125	41	1161	3806	36	900	4176	28	385	3780	20	111	2554	23	206	1560
								CAS(16,16)										
0	4	52	223	3	46	169	3	22	186	3	18	280	3	18	420	3	21	176
20	3	32	173	3	36	170	3	16	158	3	15	258	2	12	275	3	16	158
40	4	48	208	3	38	162	3	16	155	3	15	238	3	15	362	3	17	156
60	3	32	135	3	36	134	3	16	127	3	15	204	3	15	332	3	17	130
80	3	31	119	3	32	112	3	16	114	3	15	180	3	15	306	3	17	117
100	4	37	136	3	34	113	3	19	119	3	15	173	3	15	299	3	18	119
SUM	21	232	994	18	222	860	18	105	859	18	93	1333	17	90	1994	18	106	856

shown). In all cases, the orbital optimization requires quite little time.

We note that the DF calculation of the integrals does not use symmetry; the integrals are only transformed to the symmetry adapted basis at the end. Furthermore, the evaluation of the Coulomb integrals $\mathbf{J}^{kl}$ is still rather slow and not well parallelized. We hope to improve this in the future. The number of integral transformations is one more than the number of macroiterations (since a final transformation is done after the last iteration). It should also be noted that about half of the CI time is spent in the initial CI calculations and the internal optimization of each macroiteration. Compared to the CI steps, the time for density matrix evaluations is rather short since the number of density evaluations is much smaller than the number of CI steps. Furthermore, computing the 2-RDMs takes somewhat less time than the evaluation of **Hc** because for the expectation 2-RDM, one-particle transition density matrices are calculated only once, whereas for **Hc** they are needed twice.[94,95]

D. Further examples

Some further examples for CASSCF optimizations are presented in Table V. The geometries and final energies of all molecules can be found in the supplementary material. Again, the starting

TABLE IV. Elapsed times (in seconds)[a] for the $[(NH_3)_3Cu]_2O_2^{2+}$ complex using the aug-cc-pVTZ basis set and different active spaces. C_{2h} symmetry was used. "Trans." is the time for the density-fitted integral evaluation and transformation, "CI+Dens." is the time for the CI work (including density matrix evaluation), and 'Orbital Opt" is the time for the remainder.

Active space	SDs	Iterations	Transf.	CI+Dens.	Orbital Opt.	Total
CAS(8,10)[b]	11 076	4	1112	2	119	1233
CAS(16,14)[b]	2 255 121	3	894	216	98	1208
CAS(16,14)[c]	2 255 121	3	902	391	100	1393
CAS(16,16)[b]	41 410 450	3	927	4791	104	5822

[a]On a machine with Xeon E5-2650 v4@2.20GHz processors, using 6 cores.
[b]Using the QN method.
[c]Using the fully coupled method.

TABLE V. Macroiterations (Mac), orbital steps (Orb), and CI work (CI) for various CASSCF calculations.

	WMK			Uncoupled (CI)			Uncoupled (Orb)			Partially Coupled			Fully Coupled			QN-Coupled		
Molecule (state)	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI	Mac	Orb	CI
Co_2 μ-$C_2H_2(CO)_6$	7	102	390	7	182	404	7	62	472	7	48	738	7	43	1128	7	65	529
$FeCl_3(^6A_g)$[a]	5	65	174	4	59	132	4	30	152	4	23	214	4	22	364	4	31	146
$FeCl_3(^6A_g)$[b]	7	92	344	4	45	227	4	38	236	4	29	345	4	24	504	4	37	270
Al_4O_2	9	120	377	5	82	345	5	52	424	5	47	540	4	32	725	5	52	354

[a]Using AVAS threshold 0.3, target orbitals Fe(3d, 4d), active space CAS(9,12).
[b]Using AVAS threshold 0.1, target orbitals Fe(3d, 4d), active space CAS(15,15).

orbitals have been generated using the AVAS procedure. In the case of the Co_2 μ-$C_2H_2(CO)_6$ complex[127] (basis aug-cc-pVTZ), the target orbitals were Co($3d_{x^2-y^2}$, $3d_{xy}$, $3d_{xz}$, $3d_{yz}$) and for C_2 the C($2p_y$, $2p_z$) orbitals (the Co–Co bond atoms lie in the z-direction and the C–C bond in the x-direction, C_{2v} symmetry). This produced an (18,13) active space. For the planar $FeCl_3$ molecule (basis cc-pV(T+d)Z for Cl and cc-pVTZ for Fe), we computed the $^6A_1'$ state (D_{3h} symmetry). The AVAS target orbitals were Fe(3d, 4d), and selection thresholds of 0.3 and 0.1 were used. These produced CAS(9,12) and CAS(15,15) active spaces, respectively. In the latter case, the three additional active orbitals have natural occupation numbers above 1.99; we observed that also in other molecules a low AVAS threshold often leads to the inclusion of nearly doubly occupied orbitals, and this is why we mostly used a threshold of 0.3.

The last example, Al_4O_2 (aug-cc-pVDZ basis), was taken from the recent paper of Sun *et al.*,[44] in which they studied the convergence of their CASSCF method for this system. The molecule has C_s symmetry, and the active space is CAS(12,12) ($8a'$, $4a''$). The AVAS starting orbitals were obtained with the 6 oxygen 2p and the $Al_1(3p_z)$, $Al_{2,3}(3p_x,3p_z)$ atomic orbitals as targets. Using an AVAS threshold of 0.4, the desired CAS(12,12) active space was generated, and with this starting guess, the energy reported by Sun *et al.*[128] was reproduced to within $10^{-7}E_h$. The iteration numbers in Table V refer to this calculation. However, this solution is not unique. If the $Al_1(3p_y)$ orbital was added to the AVAS target space, we obtained a CASSCF solution that is 17 mE_h lower in energy, still using the same CAS(12,12) active space. This example shows that sometimes local energy minima can be found due to the competition of different correlation effects, and probably, a larger active space would be needed to avoid this problem. However, we did not investigate this further.

In all cases, the new methods clearly outperform the WMK method regarding the number of orbital and CI steps. In the calculations for $FeCl_3$ and Al_4O_2, this also affects the number of macroiterations, indicating insufficient convergence of the microiterations in the WMK method. Comparing the different algorithms, the conclusions are similar to the previous examples: the partially and fully coupled methods require many more CI steps and density evaluations than the uncoupled variants, but in these cases, the extra effort is not compensated by other savings. For example, the fully coupled calculation for Al_4O_2 needed only 4 macroiterations to converge but was nevertheless more expensive than the uncoupled ones, which needed 5 iterations. The calculations of Sun *et al.*[44] required 7 macroiterations, but since the starting orbitals were different, a direct comparison is not possible.

In all calculations of Table V, the "Uncoupled (Orb)" and QN-coupled schemes were most efficient. The QN-coupled calculations mostly needed some more CI steps than the "Uncoupled (Orb)" scheme, but nevertheless they were still slightly faster due to the simplifications in the orbital optimization (cf. Sec. V). The calculations for the cobalt complex converged rather slowly in the initial macroiterations. We have tested also other active spaces, but the convergence was always similar. Also tightening the convergence criteria in the microiterations and in the internal optimization did not lead to substantial improvements. Thus, it appears that the second-order energy approximation is less accurate in this case.

Figure 3 shows the convergence of the energies in the macroiterations for three typical cases (using the QN scheme in the microiterations). For the copper complex and $FeCl_3$, near quadratic

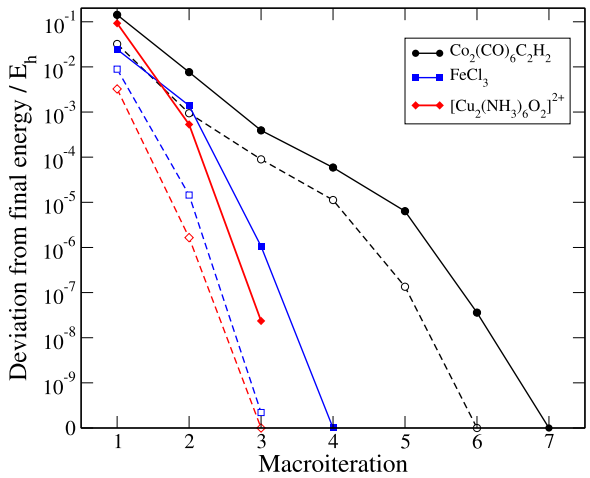

FIG. 3. Convergence of the energies in the macroiterations for $[Cu_2(NH_3)_6O_2]^{2+}$ CAS(16,14), $FeCl_3$ CAS(15,15), and $Co_2(CO)_6C_2H_2$ CAS(18,13). The full lines represent the deviations of the variational energies from the final energies after the internal optimization of each iteration. The dashed lines are the deviations of the second-order approximations $E^{(2)}(\mathbf{T}, \mathbf{c})$ at the end of each iteration. Zero deviation means that the energies differ less than 10^{-10} from the final ones.

convergence is reached already in the first and second iterations, respectively, and in these cases the convergence is extremely fast. The cobalt complex converges in the beginning more slowly and reaches the quadratic convergence region only in the fifth iteration. Most calculations presented in this paper converged in 3–4 macroiterations, similar to the $FeCl_3$ case. In the last macroiteration, the energy usually changes by less than 10^{-8} E_h. However, this iteration cannot be saved since the previous energy change or gradient is mostly above the convergence thresholds.

VII. CONCLUSIONS

We have presented a new implementation of the MCSCF optimization as originally proposed by Werner, Meyer, and Knowles.[17–20] The method differs from the previous implementation[19,20] by improved algorithms for minimizing the second-order energy approximation in the microiterations. Traditionally, this is done by an alternating (uncoupled) optimization of the orbitals and CI coefficients. Minimization of the second-order energy approximation with respect to the orbitals requires solving a system of nonlinear equations. In our new program, this is much improved by an iterative subspace method, which explicitly constructs and uses the orbital Hessian for the most critical orbital rotations. Two variants of the uncoupled optimizer are compared, one in which the focus is on minimizing the number of orbital steps and another in which the number of CI steps is minimized.

Next, a new quadratically convergent optimizer is described, which explicitly includes the coupling between the orbital rotations and the CI coefficients. Two variants of this solver are compared: in the partially coupled method, the coupling only includes the most important configurations (P-space) and one Q-space vector for each state with all orbitals. In the fully coupled method, the number of Q-space vectors is iteratively increased, and in this way, the coupling with the full configuration space can be further approached.

The disadvantage of the fully coupled method is that many additional CI steps and transition density matrix evaluations are necessary. This can be avoided by a quasi-Newton algorithm, which accounts for the orbital-CI coupling approximately at virtually no additional cost.

The convergence of the five algorithms has been demonstrated and compared for a set of 21 aromatic molecules as well as for five metal complexes. Based on these calculations and many other tests, the following conclusions can be reached:

1. In simple cases with good starting orbitals and a properly chosen active space, all methods perform almost equally well. It is then not cost effective to use the partially or fully coupled methods. This applies, for example, to the calculations of the aromatic molecules.
2. In difficult cases, where the active orbitals change character during the optimization or if the energy depends very weakly on some orbital rotations, the new orbital optimization algorithm strongly improves the convergence of the microiterations. This can also lead to improved convergence of the macroiterations since the second-order energy approximation is optimized more accurately.
3. With the uncoupled optimization, the most stable convergence is usually achieved if in each microiteration the CI optimization is done rather accurately. This increases the total number of CI steps, but, unless very large active spaces are used, the additional effort is justified by the more robust optimization. For cases with very large active spaces, in which the CI work strongly dominates, we also provide an algorithm that minimizes the number of CI iterations. However, in some cases, this can lead to an increase in the number of macroiterations.
4. When the active orbitals change character during the optimization, the coupling between orbitals and CI coefficients can become strong, leading to convergence problems in the WMK and (to a lesser extent) in the new uncoupled methods. These convergence problems disappear or are at least greatly reduced when the fully coupled method is used in the microiterations. In such difficult cases, the increase in CI work is often compensated by the overall much faster convergence.
5. The QN approximation is efficient and performs well under most conditions. It is therefore best suitable as a default method. The fully coupled method can be used as a "failsafe" option.

Despite the many tests that we performed, these conclusions are of course drawn from a limited number of applications, and future improvements may be possible. In MCSCF or CASSCF calculations, the convergence is strongly dependent on the molecule, the starting orbitals used, the active space chosen, and the number of states included in state-averaged calculations, and this makes it quite difficult to determine an optimization strategy that works equally well in all cases.

Of particular importance is a good starting guess for the orbitals. Canonical Hartree-Fock orbitals often provide poor starting orbitals, in particular, in larger molecules where the active space is rather localized. One can try to localize and/or manually reorder the HF orbitals so that the initial active space looks reasonable, but this is very cumbersome and time consuming. In the current calculations, we determined the starting orbitals using the AVAS procedure,[70] which we found extremely helpful. The technique constructs active molecular orbitals emerging from a target set of atomic valence orbitals (e.g., the metal d orbitals in a coordination complex or p_π orbitals in aromatics). It provides not only good starting orbitals but also a physically reasonable active space.

Despite the improvements achieved in the current work, CASSCF optimization remains a challenging problem. Even the best available optimization strategies may require several hundred CI steps (**Hc** and density calculations) per state. This would make very large CASSCF calculations with the DMRG or FCIQMC approaches extremely expensive. It should be noted, however, that the current methods are designed for obtaining very well converged orbitals and energies, for use in subsequent treatments of the dynamical correlation. It may be possible to obtain less converged solutions with an accuracy of, e.g., 0.1 mH in the energies with a much lower number of CI steps. With our method this accuracy is usually obtained in 2–3 macroiterations.

Even though this paper has been focused on CASSCF optimizations, the described optimization methods are also applicable to more general wavefunctions, such as restricted SCF (RASSCF)[129]

or entirely general MCSCF. Such optimizations may be more difficult than CASSCF ones since active-active orbital rotations are no longer redundant. This will be investigated in future work.

For large molecules with very many inactive orbitals, the pure second-order method as described in this work becomes unsuitable since the integral transformations become too expensive. This is also true for density fitting or Cholesky decomposition approaches since then the computation of the Coulomb integrals $J^{kl}_{rs} = (rs|ij)$ requires the computation of 3-index integrals $(rs|A)$, which scales as $\mathcal{O}(N^4_{\mathrm{AO}})$ (assuming that the number of fitting functions A is proportional to N_{AO}). This problem can be avoided by embedding the active space and its close environment in a frozen set of localized HF orbitals.[91] Then, the integrals are only required for a small set of optimized orbitals. Another approach is to use a first-order method for optimizing the inactive orbitals and a second-order method for the active space. It is then sufficient to compute in each macroiteration the integrals J^{tu}_{rs} and K^{tu}_{rs}, where t, u refer to active orbitals. However, a Fock matrix needs to be computed in each microiteration. Such an approach will be presented in Paper II.[130]

SUPPLEMENTARY MATERIAL

See supplementary material for pseudocodes of the algorithms including default parameters. Additionally, the geometries and energies of the systems presented in Sec. VI D are shown.

ACKNOWLEDGMENTS

We thank Dr. Qiming Sun for providing the structure and the energy of the Al_4O_2 cluster from Ref. 44.

APPENDIX: EVALUATION OF THE COUPLING TERMS

If the simplifications of the density matrices for closed-shell orbitals are taken into account, Eq. (51) can be rewritten as

$$\tilde{A}^{pm}_{ri} = 2\delta_{pm}P_{ri} + \sum_{tu} D^{pm}_{tu}Q^{tu}_{ri}, \tag{A1}$$

$$\tilde{A}^{pm}_{rt} = \sum_{u} P_{ru}D^{pm}_{ut} + \sum_{uvw} D^{pm}_{tu,vw}Q^{vw}_{ru}, \tag{A2}$$

where the intermediates are defined as

$$P_{rk} = (\mathbf{U}^\dagger\mathbf{F}^c\mathbf{U})_{rk} + \sum_j (\mathbf{U}^\dagger\mathbf{L}^{kj}\mathbf{T})_{rj}, \tag{A3}$$

$$Q^{vw}_{ri} = 2(\mathbf{U}^\dagger\mathbf{J}^{vw}\mathbf{U})_{ri} - (\mathbf{U}^\dagger\mathbf{K}^{vw}\mathbf{U})_{ri} + (\mathbf{U}^\dagger\mathbf{L}^{iw}\mathbf{T})_{rv}, \tag{A4}$$

$$Q^{vw}_{ru} = (\mathbf{U}^\dagger\mathbf{J}^{vw}\mathbf{U})_{ru} + 2(\mathbf{U}^\dagger\mathbf{K}^{uw}\mathbf{T})_{rv}. \tag{A5}$$

The intermediates can be used to evaluate the transformed integrals in Eqs. (55) and (56)

$$\widetilde{F}^c_{tu} = \frac{1}{2}(1+\tau_{tu})\Big[\sum_i (\Delta\mathbf{R}^\dagger\mathbf{Q}^{tu})_{ii} + (\Delta\mathbf{R}^\dagger\mathbf{P})_{tu}\Big], \tag{A6}$$

$$\widetilde{(tu|vw)} = \frac{1}{8}(1+\tau_{tu})(1+\tau_{vw})\Big[(\Delta\mathbf{R}^\dagger\mathbf{Q}^{vw})_{tu} + (\Delta\mathbf{R}^\dagger\mathbf{Q}^{tu})_{vw}\Big]. \tag{A7}$$

REFERENCES

[1] J. Hinze and C. C. J. Roothaan, Prog. Theor. Phys. Suppl. **40**, 37 (1967).
[2] B. Levy and G. Berthier, Int. J. Quantum Chem. **2**, 307 (1968).
[3] G. Das and A. C. Wahl, J. Chem. Phys. **56**, 1769 (1972).
[4] F. Grein and T. C. Chang, Chem. Phys. Lett. **12**, 44 (1971).
[5] K. K. Docken and J. Hinze, J. Chem. Phys. **57**, 4928 (1972).
[6] K. Hirao and H. Nakatsuji, J. Chem. Phys. **59**, 1457 (1973).
[7] F. Grein and A. Banerjee, Int. J. Quantum Chem. **9**, 147 (1975).
[8] A. Banerjee and F. Grein, Int. J. Quantum Chem. **10**, 123 (1976).
[9] K. Ruedenberg, L. M. Cheung, and S. T. Elbert, Int. J. Quantum Chem. **16**, 1069 (1979).
[10] B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, Chem. Phys. **48**, 157 (1980).
[11] B. O. Roos, Int. J. Quantum Chem. **18**, 175 (1980).
[12] R. Shepard and J. Simons, Int. J. Quantum Chem. **18**, 211 (1980).
[13] D. L. Yeager and P. Jørgensen, J. Chem. Phys. **71**, 755 (1979).
[14] P. E. M. Siegbahn, J. Almlöf, A. Heiberg, and B. O. Roos, J. Chem. Phys. **74**, 2384 (1981).
[15] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, and S. Elbert, Chem. Phys. **71**, 41 (1982).
[16] E. Dalgaard, Chem. Phys. Lett. **65**, 559 (1979).
[17] H.-J. Werner and W. Meyer, J. Chem. Phys. **73**, 2342 (1980).
[18] H.-J. Werner and W. Meyer, J. Chem. Phys. **74**, 5794 (1981).
[19] H.-J. Werner and P. J. Knowles, J. Chem. Phys. **82**, 5053 (1985).
[20] P. J. Knowles and H.-J. Werner, Chem. Phys. Lett. **115**, 259 (1985).
[21] H.-J. Werner, *Advances in Chemical Physics* (John Wiley & Sons, 1987), Vol. 69, pp. 1–62.
[22] D. L. Yeager and P. Jørgensen, Mol. Phys. **39**, 587 (1980).
[23] D. L. Yeager, P. Albertsen, and P. Jørgensen, J. Chem. Phys. **73**, 2811 (1980).
[24] P. Jørgensen, P. Albertsen, and D. L. Yeager, J. Chem. Phys. **72**, 6466 (1980).
[25] P. Jørgensen, P. Albertsen, and D. L. Yeager, J. Chem. Phys. **73**, 5408 (1980).
[26] P. Jørgensen, J. Olsen, and D. L. Yeager, J. Chem. Phys. **75**, 5802 (1981).
[27] A. Igawa, D. L. Yeager, and H. Fukutome, J. Chem. Phys. **76**, 5388 (1982).
[28] J. Olsen and P. Jørgensen, J. Chem. Phys. **77**, 6109 (1982).
[29] J. Olsen, P. Jørgensen, and D. L. Yeager, J. Chem. Phys. **76**, 527 (1982).
[30] J. Olsen, P. Jørgensen, and D. L. Yeager, J. Chem. Phys. **77**, 356 (1982).
[31] D. L. Yeager, D. Lynch, J. Nichols, P. Jørgensen, and J. Olsen, J. Phys. Chem. **86**, 2140 (1982).
[32] J. Olsen, D. L. Yeager, and P. Jørgensen, *Advances in Chemical Physics* (John Wiley & Sons, 1983), Vol. 54, pp. 1–176.
[33] P. Jørgensen, P. Swanstrøm, and D. L. Yeager, J. Chem. Phys. **78**, 347 (1983).
[34] B. H. Lengsfield III and B. Liu, J. Chem. Phys. **75**, 478 (1981).
[35] B. H. Lengsfield III, J. Chem. Phys. **73**, 382 (1980).
[36] B. H. Lengsfield III, J. Chem. Phys. **77**, 4073 (1982).
[37] H. J. Jensen, H. Agren, and H. Ågren, Chem. Phys. Lett. **110**, 140 (1984).
[38] H. J. Jensen and P. Jørgensen, J. Chem. Phys. **80**, 1204 (1984).
[39] H. J. Jensen and H. Ågren, Chem. Phys. **104**, 229 (1986).
[40] H. J. Jensen, P. Jørgensen, and H. Ågren, J. Chem. Phys. **87**, 451 (1987).
[41] R. Shepard, *Advances in Chemical Physics* (John Wiley & Sons, 1987), Vol. 69, pp. 63–200.
[42] F. Aquilante, T. B. Pedersen, R. Lindh, B. O. Roos, A. Sánchez de Merás, and H. Koch, J. Chem. Phys. **129**, 024113 (2008).
[43] E. G. Hohenstein, N. Luehr, I. S. Ufimtsev, and T. J. Martínez, J. Chem. Phys. **142**, 224103 (2015).
[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]K. Andersson, P.-Å. Malmqvist, and B. O. Roos, J. Chem. Phys. **96**, 1218 (1992).
[50]H.-J. Werner, Mol. Phys. **89**, 645 (1996).
[51]P. Celani and H.-J. Werner, J. Chem. Phys. **112**, 5546 (2000).
[52]C. Angeli, M. Pastore, and R. Cimiraglia, Theor. Chem. Acc. **117**, 743 (2007).
[53]P. E. Siegbahn, J. Chem. Phys. **72**, 1647 (1980).
[54]V. R. Saunders and J. H. van Lenthe, Mol. Phys. **48**, 923 (1983).
[55]H.-J. Werner and P. J. Knowles, J. Chem. Phys. **89**, 5803 (1988).
[56]P. J. Knowles and H.-J. Werner, Chem. Phys. Lett. **145**, 514 (1988).
[57]H. Werner and P. J. Knowles, Theor. Chem. Acc. **78**, 175 (1990).
[58]P. J. Knowles and H.-J. Werner, Theor. Chem. Acc. **84**, 95 (1992).
[59]P. Celani, H. Stoll, H. Werner, and P. J. Knowles, Mol. Phys. **102**, 2369 (2004).
[60]K. R. Shamasundar, G. Knizia, and H.-J. Werner, J. Chem. Phys. **135**, 054101 (2011).
[61]P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka, and R. Shepard, Chem. Rev. **112**, 108 (2012).
[62]F. A. Evangelista, M. Hanauer, A. Köhn, and J. Gauss, J. Chem. Phys. **136**, 204108 (2012).
[63]M. Hanauer and A. Köhn, Chem. Phys. **401**, 50 (2012).
[64]M. Hanauer and A. Köhn, J. Chem. Phys. **137**, 131103 (2012).
[65]M. Hanauer and A. Köhn, J. Chem. Phys. **136**, 204107 (2012).
[66]B. O. Roos, *Advances in Chemical Physics* (John Wiley & Sons, 1987), Vol. 69, pp. 399–445.
[67]J. Olsen, Int. J. Quantum Chem. **111**, 3267 (2011).
[68]G. Knizia, J. Chem. Theory Comput. **9**, 4834 (2013).
[69]C. J. Stein and M. Reiher, J. Chem. Theory Comput. **12**, 1760 (2016).
[70]E. R. Sayfutyarova, Q. Sun, G. K.-L. Chan, and G. Knizia, J. Chem. Theory Comput. **13**, 4063 (2017).
[71]J. J. Bao, S. S. Dong, L. Gagliardi, and D. G. Truhlar, J. Chem. Theory Comput. **14**, 2017 (2018).
[72]K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, and W. A. de Jong, J. Chem. Phys. **147**, 184111 (2017).
[73]G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 054106 (2009).
[74]R. E. Thomas, Q. Sun, A. Alavi, and G. H. Booth, J. Chem. Theory Comput. **11**, 5316 (2015).
[75]G. Li Manni, S. D. Smart, and A. Alavi, J. Chem. Theory Comput. **12**, 1245 (2016).
[76]J. E. Smith, B. Mussard, A. A. Holmes, and S. Sharma, J. Chem. Theory Comput. **13**, 5468 (2017).
[77]A. A. Holmes, C. J. Umrigar, and S. Sharma, J. Chem. Phys. **147**, 164111 (2017).
[78]P. M. Zimmerman, J. Chem. Phys. **146**, 104102 (2017).
[79]G. K.-L. Chan and S. Sharma, Annu. Rev. Phys. Chem. **62**, 465 (2011).
[80]D. Zgid and M. Nooijen, J. Chem. Phys. **128**, 144116 (2008).
[81]D. Ghosh, J. Hachmann, T. Yanai, and G. K.-L. Chan, J. Chem. Phys. **128**, 144117 (2008).
[82]T. Yanai, Y. Kurashige, D. Ghosh, and G. K.-L. Chan, Int. J. Quantum Chem. **109**, 2178 (2009).
[83]Y. Ma and H. Ma, J. Chem. Phys. **138**, 224105 (2013).
[84]D. R. Yarkony, Chem. Phys. Lett. **77**, 634 (1981).
[85]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 242 (2012).
[86]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz *et al.*, MOLPRO, version 2019.1, a package of *ab initio* programs, 2019, see http://www.molpro.net.
[87]J. W. Snyder, E. G. Hohenstein, N. Luehr, and T. J. Martínez, J. Chem. Phys. **143**, 154107 (2015).
[88]B. S. Fales and B. G. Levine, J. Chem. Theory Comput. **11**, 4708 (2015).
[89]J. W. Snyder, B. F. E. Curchod, and T. J. Martínez, J. Phys. Chem. Lett. **7**, 2444 (2016).
[90]J. W. Snyder, B. S. Fales, E. G. Hohenstein, B. G. Levine, and T. J. Martínez, J. Chem. Phys. **146**, 174113 (2017).
[91]F. Menezes, D. Kats, and H.-J. Werner, J. Chem. Phys. **145**, 124115 (2016).
[92]R. Shepard, Theor. Chem. Acc. **84**, 55 (1992).
[93]E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).
[94]P. E. M. Siegbahn, Chem. Phys. Lett. **109**, 417 (1984).
[95]P. J. Knowles and N. C. Handy, Chem. Phys. Lett. **111**, 315 (1984).
[96](a) C. G. Broyden, IMA J. Appl. Math. **6**, 76 (1970); (b) R. Fletcher, Comput. J. **13**, 317 (1970); (c) D. Goldfarb, Math. Comp. **24**, 23 (1970); (d) D. F. Shanno, *ibid.* **24**, 647 (1970).
[97]R. Fletcher, *Practical Methods of Optimization* (John Wiley & Sons, 2013).
[98]M. W. Feyereisen, G. Fitzgerald, and A. Komornicki, Chem. Phys. Lett. **208**, 359 (1993).
[99]W. Győrffy, T. Shiozaki, G. Knizia, and H.-J. Werner, J. Chem. Phys. **138**, 104104 (2013).
[100]R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[101]S. F. Sontum, D. A. Case, and M. Karplus, J. Chem. Phys. **79**, 2881 (1983).
[102]C. Rovira, K. Kunc, J. Hutter, P. Ballone, and M. Parrinello, J. Phys. Chem. A **101**, 8914 (1997).
[103]B. Delley, Physica B **172**, 185 (1991).
[104]R. Olivares-Amaya, W. Hu, N. Nakatani, S. Sharma, J. Yang, and G. K.-L. Chan, J. Chem. Phys. **142**, 034102 (2015).
[105]H. Chen, W. Lai, and S. Shaik, J. Phys. Chem. B **115**, 1727 (2011).
[106]S. Vancoillie, H. Zhao, V. T. Tran, M. F. A. Hendrickx, and K. Pierloot, J. Chem. Theory Comput. **7**, 3961 (2011).
[107]M.-S. Liao, J. D. Watts, and M.-J. Huang, J. Comput. Chem. **27**, 1577 (2006).
[108]W. D. Edwards, B. Weiner, and M. C. Zerner, J. Am. Chem. Soc. **108**, 2196 (1986).
[109]Y.-K. Choe, T. Nakajima, K. Hirao, and R. Lindh, J. Chem. Phys. **111**, 3837 (1999).
[110]A. R. Groenhof, M. Swart, A. W. Ehlers, and K. Lammertsma, J. Phys. Chem. A **109**, 3411 (2005).
[111]K. Pierloot, Mol. Phys. **101**, 2083 (2003).
[112]Q. M. Phung, S. Wouters, and K. Pierloot, J. Chem. Theory Comput. **12**, 4352 (2016).
[113]K. Pierloot, Q. M. Phung, and A. Domingo, J. Chem. Theory Comput. **13**, 537 (2017).
[114]Q. M. Phung, M. Feldt, J. N. Harvey, and K. Pierloot, J. Chem. Theory Comput. **14**, 2446 (2018).
[115]G. Li Manni and A. Alavi, J. Phys. Chem. A **122**, 4935 (2018).
[116]G. Li Manni, D. Kats, D. P. Tew, and A. Alavi, J. Chem. Theory Comput. **15**, 1492 (2019).
[117]T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[118]N. B. Balabanov and K. A. Peterson, J. Chem. Phys. **123**, 064107 (2005).
[119]B. F. Gherman and C. J. Cramer, Coord. Chem. Rev. **253**, 723 (2009).
[120]M. Flock and K. Pierloot, J. Phys. Chem. A **103**, 95 (1999).
[121]M. F. Rode and H.-J. Werner, Theor. Chem. Acc. **114**, 309 (2005).
[122]C. J. Cramer, M. Włoch, P. Piecuch, C. Puzzarini, and L. Gagliardi, J. Phys. Chem. A **110**, 1991 (2006).
[123]P. Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer, and L. Gagliardi, J. Chem. Phys. **128**, 204109 (2008).
[124]K. H. Marti, I. M. Ondík, G. Moritz, and M. Reiher, J. Chem. Phys. **128**, 014104 (2008).
[125]T. Yanai, Y. Kurashige, E. Neuscamman, and G. K.-L. Chan, J. Chem. Phys. **132**, 024105 (2010).
[126]G. Barcza, Ö. Legeza, K. H. Marti, and M. Reiher, Phys. Rev. A **83**, 012508 (2011).
[127]J. A. Platts, G. J. S. Evans, M. P. Coogan, and J. Overgaard, Inorg. Chem. **46**, 6291 (2007).
[128]Q. Sun, private communication (2019).
[129]P.-Å. Malmqvist, A. Rendell, and B. O. Roos, J. Phys. Chem. **94**, 5477 (1990).
[130]D. A. Kreplin, P. J. Knowles, and H.-J. Werner, "Second-order MCSCF optimization revisited. II. Combined first- and second-order optimization algorithms for large molecules," J. Chem. Phys. (to be published).