

A second order multiconfiguration SCF procedure with optimum convergence

HansJoachim Werner and Peter J. Knowles

Citation: *J. Chem. Phys.* **82**, 5053 (1985); doi: 10.1063/1.448627

View online: <http://dx.doi.org/10.1063/1.448627>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v82/i11>

Published by the [American Institute of Physics](#).

Additional information on J. Chem. Phys.

Journal Homepage: <http://jcp.aip.org/>

Journal Information: http://jcp.aip.org/about/about_the_journal

Top downloads: http://jcp.aip.org/features/most_downloaded

Information for Authors: <http://jcp.aip.org/authors>

ADVERTISEMENT

physicstoday

Comment on any
Physics Today article.

The advertisement shows a red arrow pointing from the text 'Comment on any Physics Today article.' to a comment box on a sample article. The sample article is titled 'Measured energy in Japan' by David von Seggern. The comment box contains a comment by Edgar McCarroll dated 14 July 2012 19:59, discussing the energy release of a 100-megaton explosion and the damage potential of nuclear bombs.

A second order multiconfiguration SCF procedure with optimum convergence

Hans-Joachim Werner^{a)} and Peter J. Knowles

University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, United Kingdom

(Received 23 October 1984; accepted 23 November 1984)

An MCSCF procedure is described which is based on the direct minimization of an approximate energy expression which is periodic and correct to second order in the changes in the orthonormal orbitals. Within this approximation, the CI coefficients are fully optimized, thereby accounting for the coupling between orbital rotations and CI coefficients to higher order than in previous treatments. Additional transformations among the internal orbitals and their associated one- and two-electron integrals are performed which amounts to treating the rotations among internal orbitals to higher than second order. These extra steps are cheap compared to the four index transformation performed in each iteration, but lead to a remarkable enhancement of convergence and overall efficiency. In all calculations attempted to date, convergence has been achieved in at most three iterations. The energy has been observed to converge better than quadratically from the first iteration even when the initial Hessian matrix has many negative eigenvalues.

I. INTRODUCTION

Second order multiconfiguration self-consistent field (MCSCF) methods¹⁻⁹ or approximate second order MCSCF methods¹⁰⁻¹⁴ today occupy a central place among the tools available for performing molecular electronic structure calculations. Recently, "direct" second order MCSCF procedures have become available,^{6,8,9,15} which avoid the explicit construction of the Hamiltonian and Hessian matrices, and therefore make possible the full optimization of quite long CI expansions. Furthermore, it is now possible to perform accurate multireference configuration interaction calculations at rather modest expense,¹⁶⁻²² and these calculations usually require MCSCF wave functions as an initial approximation. It is therefore essential to be able to obtain MCSCF wave functions reliably and efficiently.

Most modern MCSCF procedures are based on exponential transformations $U = \exp(\mathbf{R})$ of the orbitals, where \mathbf{R} is an antisymmetric matrix of independent variational parameters.^{23,24} The simplest approach based on this formulation is to approximate the energy E as a second order function of \mathbf{R} and the changes in the CI coefficients Δc . Minimization of this approximate functional yields a system of linear equations for the parameters R_{ri} and Δc_i . This Newton-Raphson (NR) method converges quadratically when near to the solution. However, it has been found by experience to have a very small radius of convergence. Many damping schemes have therefore been suggested to improve the global convergence of the NR method.^{3,7,15,25-31} Possibly the most effective and inexpensive of these is the Augmented Hessian (AH) method, which was originally proposed by Lengsfeld,³ and is now used by many workers in various modifications.^{6,7,9,30,31} In this method, the linear equations are

replaced by an eigenvalue equation which automatically introduces a level shift into the Hessian. Furthermore, the eigenvalue equation can be damped with a single parameter, which has the effect of changing the level shift.^{9,30} This parameter is chosen such that the norm of the solution is smaller than some automatically updated value, and convergence is then guaranteed as shown by Fletcher,³² and extensively discussed by Jørgensen *et al.*^{9,30,31} This of course does not mean that convergence will be reached in an acceptable number of iterations. In examples of average complexity given recently by Jørgensen *et al.*,^{15,30,31} to reach the desired solution up to ten iterations were required, of which only the last two or three were in the quadratic local region. This is also in line with our own experience using this method. Because the computational effort for each MCSCF iteration is often considerable, it is desirable to examine whether it is possible to find an alternative method which takes into account more information about the true energy surface than the NR method, and therefore converges more quickly to the local region. A straightforward extension of the NR method would be to minimize an energy approximation correct to third or higher order in \mathbf{R} ,^{15,33,34} but this requires a much more expensive integral transformation in each iteration than a second order method.

One of the reasons for the small radius of convergence of the NR method is the fact that the true energy is a periodic rather than quadratic function of the orbital rotation generator \mathbf{R} . When far from the desired solution, the matrix of second derivatives with respect to the variational parameters (Hessian matrix) therefore often has many negative eigenvalues, and the approximate function $E(R^2)$ does not properly represent the true energy. This phenomenon was recognized by Werner and Meyer,^{4,5} who proposed the use of an energy functional $E(T^2)$ correct to second order in $\mathbf{T} = \mathbf{U} - \mathbf{1}$. Since in this approximation the orthonormality condition on the orbitals is taken into account exactly, it is periodic in the

^{a)} Permanent address: Fachbereich Chemie der Universität Kaiserslautern, D-675 Kaiserslautern, West Germany.

orbital changes, and a minimum of this approximation can always be found, even if the Hessian matrix at the expansion point is not positive definite. It has been observed in many applications that the method of Werner and Meyer (WM) has a large radius of convergence, and, particularly when the Hessian matrix is not positive definite, it usually converges to the local region much more quickly than the AH method. The fact that this approach contains certain terms of infinite order in \mathbf{R} introduces a complication in that the Newton–Raphson equations are replaced by nonlinear equations, but this has not proved to be a serious practical difficulty. Since no additional transformed integrals are required, the cost per iteration of the WM and NR methods is about the same.

However, there are two features of the WM method which we have found to be not entirely satisfactory. Firstly, when far from the solution, the inclusion of coupling between the orbital and CI coefficient changes often causes a deterioration rather than an improvement in convergence. This is probably because the coupling terms are included to only first order in the orbital changes; in the initial iterations, however, it is not uncommon for orbital rotations to be as large as 60° , and then the second order terms are expected to be as significant as the first order ones. The second undesirable property of the WM method is that it usually underestimates rotations between strongly occupied orbitals. This can lead to rather slow convergence even in the local region. The reason for this deficiency will be analyzed in this paper; it originates from the fact that the approximate functional $E(T^2)$ is not invariant to rotations between doubly occupied orbitals. Because of these two difficulties it has been our practice to leave out the CI-orbital coupling in initial iterations, and also to switch to the AH or NR method once the internal-external rotations become small. However it is not always easy to decide in advance under what conditions these switches should be made. It is therefore desirable to have a single method which works well at all stages of a calculation.

This paper is concerned with the elimination of the above problems. Firstly, it will be shown that in direct MCSCF methods, in which the Hessian matrix is never explicitly constructed, the coupling terms between orbital and CI coefficient changes can be easily accounted for to second order in both the orbital and CI coefficient changes with very little extra computational effort. In fact, for cases with many configurations the effort may even be reduced as compared to the AH or NR methods. Secondly, we will treat the rotations between internal orbitals effectively up to fourth order in \mathbf{T} , which restores the invariance of the energy to rotations between doubly occupied orbitals. The small additional computational effort to do so is easily outweighed by a drastic improvement of convergence. It will be demonstrated that our new method exhibits very much better convergence than all previously applied procedures. In fact, quadratic convergence behavior is observed from the very beginning, even in cases with more than 20 negative eigenvalues in the initial Hessian. In all cases tested so far, convergence of the

energy to better than 10^{-10} a.u. has been obtained in three iterations.

II. ORBITAL OPTIMIZATION

In this section we outline our orbital optimization procedure. For a given set of CI coefficients, this method is equivalent to the method of Werner and Meyer.^{4,5} Our method differs from the latter, however, in the optimization of the CI coefficients and the internal orbital rotations. This will be explained in Secs. IV and V.

Our aim is to minimize the energy of a wave function of the form

$$\Psi = \sum_I c_I \Phi_I \quad \text{with} \quad \sum_I c_I^2 = 1 \quad \text{and} \quad \langle \Phi_I | \Phi_J \rangle = \delta_{IJ} \quad (1)$$

with respect to the CI coefficients $\{c_I\}$ and the orbitals $\{\phi_i\}$ from which the configurations $\{\Phi_I\}$ are constructed. The orbitals are represented as linear combinations of basis functions $\{\chi_\mu\}$ (AOs)

$$\phi_i \equiv |i\rangle = \sum_\mu C_{\mu i} \chi_\mu \quad (2)$$

and are assumed to be orthonormal, i.e.,

$$\langle i | j \rangle = \int \phi_i^*(1) \phi_j(1) d\tau_1 = \delta_{ij}. \quad (3)$$

In terms of a given set of CI coefficients and orbitals, the energy takes the form

$$E_0 = \sum_{ij} \langle i | h | j \rangle \gamma_{ij} + \frac{1}{2} \sum_{ijkl} (ij | kl) \Gamma_{ijkl}, \quad (4)$$

where

$$\langle i | h | j \rangle = \int \phi_i^*(1) \hat{h} \phi_j(1) d\tau_1 \quad (5)$$

and

$$(ij | kl) = \int \phi_i^*(1) \phi_k^*(2) r_{12}^{-1} \phi_j(1) \phi_l(2) d\tau_1 d\tau_2 \quad (6)$$

are the one and two electron integrals in the MO basis.

$$\gamma_{ij} = \sum_{IJ} c_I c_J \gamma_{ij}^{IJ} \quad (7)$$

and

$$\Gamma_{ijkl} = \frac{1}{2} \sum_{IJ} c_I c_J (\Gamma_{ijkl}^{IJ} + \Gamma_{jikl}^{IJ}) \quad (8)$$

are the first and second order reduced density matrices, respectively. The coupling coefficients γ_{ij}^{IJ} and Γ_{ijkl}^{IJ} depend only on the structure of the wave function. Assuming that the wave function is real, we may restrict the orbitals to be real and the density matrices have been symmetrized such that

$$\gamma_{ij} = \gamma_{ji}, \quad \Gamma_{ijkl} = \Gamma_{jikl} = \Gamma_{klij}. \quad (9)$$

A new set of orthonormal molecular orbitals $\{\tilde{\phi}_i\}$ may be obtained by an orthogonal transformation

$$|\tilde{i}\rangle = \sum_r |r\rangle U_{ri}, \quad (10)$$

where \mathbf{U} can be represented in exponential form

$$\mathbf{U}(\mathbf{R}) = \exp(\mathbf{R}) = \mathbf{1} + \mathbf{R} + \frac{1}{2}\mathbf{R}\mathbf{R} + \dots, \quad (11)$$

with

$$\mathbf{R} = -\mathbf{R}^\dagger. \quad (12)$$

The elements R_{ri} ($r > i$) of the antisymmetric matrix \mathbf{R} constitute a set of independent variational parameters for the orbital rotations. Redundant variables R_{ij} are determined as described in Ref. 35 and are eliminated. Here and in the following we denote internal (occupied) orbitals by the labels i, j, k, l , and any orbitals (occupied and unoccupied) by the labels r, s .

If the new orbitals $|\tilde{i}\rangle$ are inserted into the energy expression (4), one obtains a fourth order function of \mathbf{U} . Since it is impractical to optimize this energy functional directly with respect to \mathbf{U} , we expand it up to second order in

$$\mathbf{T} = \mathbf{U} - \mathbf{1} \quad (13)$$

and minimize the resulting second order energy approximation $E(T^2)$. The process is repeated until convergence is reached. Note that the second order energy approximation in \mathbf{T} will contain terms of infinite order in \mathbf{R} , whereas in the Newton-Raphson or Augmented Hessian methods only terms up to second order in \mathbf{R} are considered.

Up to second order in \mathbf{T} and for a given set of CI coefficients, the energy takes the form

$$\begin{aligned} E^{(2)}(\mathbf{T}) = E_0 + 2 \sum_{ri} T_{ri} \left[\sum_j \langle r|h|j \rangle \gamma_{ij} + \sum_{jkl} \langle r|J^{kl}|j \rangle \Gamma_{ijkl} \right] \\ + \sum_{ri} \sum_{sj} T_{ri} T_{sj} \left[\langle r|h|s \rangle \gamma_{ij} + \sum_{kl} \langle r|J^{kl}|s \rangle \Gamma_{ijkl} \right] \\ + 2 \sum_{kl} \langle r|K^{kl}|s \rangle \Gamma_{ikjl}. \end{aligned} \quad (14)$$

Here we have introduced the generalized Coulomb and exchange operators

$$\langle r|J^{kl}|s \rangle = (rs|kl), \quad (15)$$

$$\langle r|K^{kl}|s \rangle = (rk|ls), \quad (16)$$

which represent the subset of the two electron integrals in the MO basis with at least two internal orbitals. These operators are required in any second order MCSCF method and have to be recalculated in each iteration. Efficient algorithms for their evaluation have been described by several authors.^{4,10,22}

The stationary condition for the energy approximation (14) with respect to \mathbf{R} is easily derived by defining an update of the unitary matrix $\mathbf{U}(\mathbf{R})$ as

$$\mathbf{U}(\mathbf{R}, \Delta\mathbf{R}) \equiv \mathbf{U}(\mathbf{R}) \cdot \mathbf{U}(\Delta\mathbf{R}) \quad (17)$$

and, for a given $\mathbf{U}(\mathbf{R})$, takes the form

$$\left(\frac{\partial E^{(2)}}{\partial \Delta R_{ri}} \right)_{\Delta\mathbf{R}=0} = (\mathbf{U}^\dagger \mathbf{B} - \mathbf{B}^\dagger \mathbf{U})_{ri} = 0. \quad (18)$$

An alternative derivation of these equations has been given in Ref. 4. The matrix \mathbf{B} is defined by

$$\begin{aligned} B_{ri} = 2 \left[\sum_{sj} \langle r|h|s \rangle U_{sj} \gamma_{ji} + \sum_{sj} \sum_{kl} \langle r|J^{kl}|s \rangle U_{sj} \Gamma_{jikl} \right. \\ \left. + 2 \sum_{sj} \sum_{kl} \langle r|K^{kl}|s \rangle T_{sj} \Gamma_{jikl} \right]. \end{aligned} \quad (19)$$

In order to obtain the optimum matrix \mathbf{U} , the nonlinear equations (18) have to be solved. This can be done iteratively by a simple relaxation scheme, i.e.,

$$\Delta R_{ri} = -(\mathbf{U}^\dagger \mathbf{B} - \mathbf{B}^\dagger \mathbf{U})_{ri} / D_{ri}, \quad (20)$$

where D_{ri} are suitable positive denominators, e.g., the diagonal second energy derivatives with respect to ΔR_{ri} at $\mathbf{R} = \mathbf{0}$. Better values D_{ri} can be obtained during the iteration using a finite difference method as described in Ref. 4. Convergence can be improved by applying the DIIS procedure devised by Pulay.³⁶ A more stable method to solve Eq. (18) is presented in Sec. III.

In the following, we denote the iterations performed to solve the nonlinear equations (18) "microiterations," as opposed to the "macroiterations" in which new sets of Coulomb and exchange operators and a new variational energy are calculated. In each microiteration, the matrix \mathbf{B} has to be evaluated. This is accomplished by performing a one index transformation of the operators \mathbf{h} , \mathbf{J}^{kl} , and \mathbf{K}^{kl} .

$$\tilde{\mathbf{h}}_{rj} = (\mathbf{h}\mathbf{U})_{rj}, \quad (21)$$

$$(\tilde{\mathbf{J}}^{kl})_{rj} = (\mathbf{J}^{kl}\mathbf{U})_{rj}, \quad (22)$$

$$(\tilde{\mathbf{K}}^{kl})_{rj} = (\mathbf{K}^{kl}\mathbf{T})_{rj}. \quad (23)$$

This transformation requires about $\frac{3}{2}N^2M^3$ operations (M = number of internal orbitals, N = total number of orbitals), and is therefore very cheap as compared to the four index transformation needed to evaluate new operators \mathbf{J}^{kl} , \mathbf{K}^{kl} (about $\frac{3}{8}N^4M + 3N^3M^2$ operations). Note that a similar one index transformation has to be performed in any direct MCSCF method.

In concluding this section we note that orbitals which are doubly occupied in all configurations can be treated without evaluating and storing the coupling coefficients and density matrix elements which involve any indices corresponding to such orbitals. Explicit formulas for this case are given in Eqs. (62)–(70) in Ref. 4. It is not recommended, however, to optimize core orbitals by a first order procedure as described in Eqs. (71)–(74) of Ref. 4, since this would certainly spoil convergence in most cases. Instead, energetically low lying core orbitals should be determined in a preceding SCF calculation and then kept frozen. This considerably reduces the computational effort for large molecules.

III. SOLUTION OF THE NONLINEAR EQUATIONS

In order to make our method useful, it is essential to have a stable and efficient algorithm to solve the nonlinear equations (18). We have found that the simple scheme outlined in the previous section is not always

entirely satisfactory, and therefore we have devised an alternative method, which works considerably better.

The method uses the idea of linearizing equation (18) by expanding the energy $E^{(2)}(\mathbf{T})$ to second order in $\Delta\mathbf{R}$ at an expansion point defined by an arbitrary trial solution \mathbf{U} . Then the Augmented Hessian approach can be used to obtain $\Delta\mathbf{R}$, and hence a new \mathbf{U} according to Eq. (17). Following the definition of Eq. (17) we can write

$$\mathbf{T}(\mathbf{R}, \Delta\mathbf{R}) = \mathbf{T} + \Delta\mathbf{T},$$

$$\mathbf{T} = \mathbf{U}(\mathbf{R}) - \mathbf{1},$$

$$\Delta\mathbf{T} = \mathbf{U}(\mathbf{R})(\Delta\mathbf{R} + \frac{1}{2}\Delta\mathbf{R}^2 + \dots), \quad \Delta\mathbf{R} = -\Delta\mathbf{R}^\dagger. \quad (24)$$

Inserting this into the energy expression (14), we obtain, to second order in $\Delta\mathbf{R}$,

$$E^{(2)}(\mathbf{T}, \Delta\mathbf{R}) = E^{(2)}(\mathbf{T}) + 2 \sum_{ri} (\Delta R_{ri} + \frac{1}{2}(\Delta\mathbf{R}^2)_{ri}) A_{ri} + \sum_{sitj} (\Delta R_{st} (\mathbf{U}^\dagger \mathbf{G}^{ij} \mathbf{U})_{st} \Delta R_{tj}), \quad (25)$$

where

$$(\mathbf{G}^{ij})_{rs} = h_{rs} \gamma_{ij} + \sum_{kl} \langle r | \mathbf{J}^{kl} | s \rangle \Gamma_{ijkl} + 2 \sum_{kl} \langle r | \mathbf{K}^{kl} | s \rangle \Gamma_{ikjl} \quad (26)$$

and

$$A_{ri} = (\mathbf{U}^\dagger \mathbf{B})_{ri}. \quad (27)$$

Defining further

$$\begin{aligned} \tilde{\mathbf{B}}_{ri} &= B_{ri} + \sum (\mathbf{G}^{ij} \mathbf{U} \Delta\mathbf{R})_{rj} \\ &= B_{ri} + \sum_j (\mathbf{h} \mathbf{U} \Delta\mathbf{R})_{rj} \gamma_{ji} + \sum_{jkl} (\mathbf{J}^{kl} \mathbf{U} \Delta\mathbf{R})_{rj} \Gamma_{jikl} \\ &\quad + 2 \sum_{jkl} (\mathbf{K}^{kl} \mathbf{U} \Delta\mathbf{R})_{rj} \Gamma_{jlik}, \end{aligned} \quad (28)$$

we obtain the following stationary condition for $\Delta\mathbf{R}$:

$$\mathbf{U}^\dagger \tilde{\mathbf{B}} - \tilde{\mathbf{B}}^\dagger \mathbf{U} - \frac{1}{2}[\mathbf{A} + \mathbf{A}^\dagger] \Delta\mathbf{R} + \Delta\mathbf{R}(\mathbf{A} + \mathbf{A}^\dagger) = \mathbf{0}. \quad (29)$$

This equation is linear in $\Delta\mathbf{R}$ and reduces to the well known Newton–Raphson equation for $\mathbf{U} = \mathbf{1}$ (cf. Ref. 4). However, instead of solving these equations directly, we apply the Augmented Hessian method with step restrictions. This has the great advantage that a reasonable solution is always found, even if the Hessian at the expansion point is not positive definite. The eigenvalue equation to be solved differs from the above linear equations by a simple level shift term:

$$\begin{aligned} \mathbf{U}^\dagger \tilde{\mathbf{B}} - \tilde{\mathbf{B}}^\dagger \mathbf{U} - \frac{1}{2}[\mathbf{A} + \mathbf{A}^\dagger] \Delta\mathbf{R} \\ + \Delta\mathbf{R}(\mathbf{A} + \mathbf{A}^\dagger) - \lambda \epsilon \Delta\mathbf{R} = \mathbf{0}, \end{aligned}$$

where

$$\epsilon = \lambda \operatorname{tr}(\mathbf{A}^\dagger \Delta\mathbf{R}). \quad (30)$$

$\lambda \geq 1$ is a damping parameter which can be adjusted automatically such that $\sum_{ri} \Delta R_{ri}^2 \leq s^2$, where s is a prescribed threshold defining the maximum step length.

The eigenvalue equation is solved iteratively using Davidson's method.³⁷ Note that the construction of $\tilde{\mathbf{B}}$ in each of these microiterations proceeds along completely similar lines to the calculation of \mathbf{B} , with the only difference being that in Eqs. (21)–(23) the matrix $\mathbf{U} \Delta\mathbf{R}$ replaces both \mathbf{U} and \mathbf{T} , and finally \mathbf{B} is added. Since the effort for obtaining \mathbf{B} and $\tilde{\mathbf{B}}$ is the same, only a few (typically 3–4) linear steps are performed between updates of \mathbf{U} [Eq. (17)] and \mathbf{A} . Each time \mathbf{U} is updated, new CI coefficients can be obtained as well. This is described in the next section.

IV. SIMULTANEOUS OPTIMIZATION OF ORBITALS AND CI COEFFICIENTS

So far, we have not considered any relaxation of the CI coefficients, and the orbital optimization scheme described in the previous section is equivalent to the method of Werner and Meyer.⁴ In the latter method, changes of the CI coefficients were allowed for by adding to the energy approximation (14) terms proportional to $\mathbf{T} \cdot \Delta c_I$ and $\Delta c_I \Delta c_J$. This results in additional terms in the matrix \mathbf{B} , which are accurate to first order in the Δc_I . The Δc_I themselves are obtained accurate to first order in \mathbf{T} .

In our new method, we treat the coupling of the orbital rotations and CI coefficients more accurately, but perhaps in an even simpler way. When evaluating the one-index-transformed integrals [Eqs. (21)–(23)] we also perform the second half transformation to the operators \mathbf{h} , \mathbf{J}^{kl} , and \mathbf{K}^{kl} , and evaluate the following one electron and two electron integrals:

$$h_{ij}^{(2)} = (\mathbf{U}^\dagger \mathbf{h} \mathbf{U})_{ij}, \quad (31)$$

$$\begin{aligned} (ij|kl)^{(2)} &= -(ij|kl) + (\mathbf{U}^\dagger \mathbf{J}^{kl} \mathbf{U})_{ij} + (\mathbf{U}^\dagger \mathbf{J}^{ij} \mathbf{U})_{kl} \\ &\quad + (1 + \tau_{ij})(1 + \tau_{kl})(\mathbf{T}^\dagger \mathbf{K}^{ik} \mathbf{T})_{jl}. \end{aligned} \quad (32)$$

The permutation operator τ_{ij} exchanges the labels i and j . These integrals, which can easily be kept in high speed memory, are accurate to second order in \mathbf{T} . Their evaluation requires about $\frac{1}{4}NM^4$ additional operations, and is therefore cheap as compared to the first half transformation [Eqs. (21)–(23)]. These integrals are then used in a direct CI step in order to improve the CI coefficients. For the case that one configuration strongly dominates the wave function, the update may be performed by first order perturbation theory, i.e.,

$$\tilde{c}_I = c_I - (g_I^{(2)} - E^{(2)} c_I) / (H_I^{(2)} - E^{(2)}), \quad (33)$$

for all but the dominant configuration. Here, $H_I^{(2)}$ are the diagonal elements of the Hamiltonian matrix accurate to second order in \mathbf{T} , and the elements $g_I^{(2)}$ are defined as

$$g_I^{(2)} = \sum_J c_J \left[\sum_{ij} h_{ij}^{(2)} \gamma_{ij}^I + \frac{1}{2} \sum_{ij} \sum_{kl} (ij|kl)^{(2)} \Gamma_{ijkl}^I \right]. \quad (34)$$

The second order energy expectation value $E^{(2)}$ is given by

$$E^{(2)} = \sum_I g_I^{(2)} c_I / \sum_I c_I^2. \quad (35)$$

For the case that no configuration strongly dominates the wave function, or the energy average of several states is to be minimized (see below), a modified scheme is employed in which part of the Hamiltonian matrix is explicitly constructed and diagonalized (see Appendix A). The new set of CI coefficients $\{c_I\}$ is finally normalized and used to evaluate new density matrices γ_{ij} and Γ_{ijkl} . These are then used, together with the transformed operators $\tilde{\mathbf{h}}, \tilde{\mathbf{J}}^{kl}, \tilde{\mathbf{K}}^{kl}$ [Eqs. (21)–(23)] to evaluate the \mathbf{B} matrix; this introduces the coupling into the formalism. Note that for a given set of CI coefficients and a given \mathbf{U} matrix the second order energy expressions (14) and (35) are identical. Hence, in our method, the energy expectation value $E^{(2)}$ in Eq. (35) is minimized with respect to \mathbf{R} and \mathbf{c} simultaneously.

At this point it should be pointed out again that the above treatment is more accurate than previous ones, since the final CI coefficients in each macroiteration are accurate to second order in \mathbf{T} and exactly normalized, and the final \mathbf{B} matrix is accurate to full order in the new \mathbf{c} . Nevertheless, the computational effort per microiteration for large CI expansions may be smaller than for a corresponding Newton–Raphson or Augmented Hessian calculation. In the latter cases, in each microiteration the vectors $\sum_I H_{n,I} \Delta c_I$, $\sum_{r,i} H_{I,ri} R_{ri}$, $\sum_j H_{I,j} \Delta c_j$, where \mathbf{H} is the Hessian matrix, have to be evaluated, which effectively requires three direct CI steps. In our present treatment, only two CI steps are needed: one for the evaluation of \mathbf{g} , and the second for calculating the new density matrices. However, two passes through the formula tape per microiteration are required, whereas in the NR or AH method one pass is sufficient. This is only a slight disadvantage, since very large CASSCF (complete active space MCSCF) calculations can be performed without storing the coupling coefficients at all. This may be accomplished using a method proposed by Siegbahn³⁸ and modified by Knowles and Handy.³⁹ A further advantage of our method is that the CI updates do not have to be performed in every microiteration. In fact, as already mentioned in Sec. III, the CI coefficients are improved only after performing typically 3–5 microiterations, in which Eq. (18) is linearized. Furthermore, the CI update is performed only when the orbital changes resulting from the solution of the Augmented Hessian equations as described in Sec. III become lower than a prescribed threshold. Although this may result in a larger total number of microiterations, the number of direct CI steps is minimized. This appears to be important in cases with many configurations, in which the direct CI steps dominate the whole computational effort.⁴²

We also note that our method allows the optimization of an energy average of several states.^{5,8,10,40,42} For this purpose, in each microiteration two or more eigenvectors $\{c_I^n\}$ are updated as described in the Appendix and used to evaluate averaged density matrices

$$\gamma_{ij} = \sum_n w_n \sum_{IJ} c_I^n c_J^n \gamma_{ij}^{IJ}, \quad (36)$$

$$\Gamma_{ijkl} = \sum_n w_n \sum_{IJ} c_I^n c_J^n \Gamma_{ijkl}^{IJ}, \quad (37)$$

where the w_n are arbitrary weight factors for the states n . These averaged density matrices then have to be used to calculate \mathbf{B} .

V. OPTIMIZATION OF THE INTERNAL-INTERNAL ORBITAL ROTATIONS

So far, our method differs from the method of Werner and Meyer mainly by the inclusion of the higher order terms coupling the orbital rotations and changes of the CI coefficients. As will be demonstrated in the next section, this considerably improves convergence in CASSCF calculations in which no rotations between internal orbitals have to be performed. However, it does not remedy the slow convergence of rotations between strongly occupied orbitals mentioned already in the introduction. In order to understand this problem we have investigated the quality of our energy approximation for simple model systems. The accuracy of various energy approximations in a three configuration MCSCF wave function for the HF molecule for different kinds of orbital rotations is demonstrated in Figs. 1 and 2. In both cases, canonical Hartree–Fock orbitals have been used as starting guess, and only one orbital rotation is performed at a time with the other R_{ij} kept zero. The CI coefficients are relaxed as in the corresponding methods (i.e., the “partitioned” Hessian matrix elements^{2,5,7,26} which include the CI coupling are used in the NR and AH cases). Figure 1 shows the energy functionals for a rotation between a weakly occupied and an unoccupied orbital. Since the second energy derivative at the expansion point is negative,

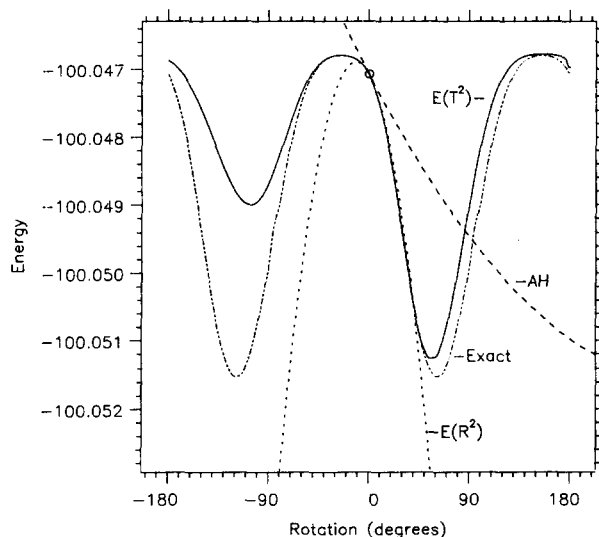


FIG. 1. The dependence of various energy approximations on the rotation $4\sigma-5\sigma$ for a 3 configuration MCSCF calculation for the HF molecule at $r = 1.6$ a.u. Basis set: Huzinaga (Ref. 43) $9s5p$ (innermost $4s$ and $2p$ contracted), augmented by a d function (exponent 1.5), on F; Huzinaga $5s$ (innermost $2s$ contracted), augmented by a p function (exponent 0.75), on H. Configurations: $1\sigma^2 2\sigma^2 3\sigma^2 1\pi^4$, $1\sigma^2 2\sigma^2 4\sigma^2 1\pi^4$, $1\sigma^2 3\sigma^2 4\sigma^2 1\pi^4$.

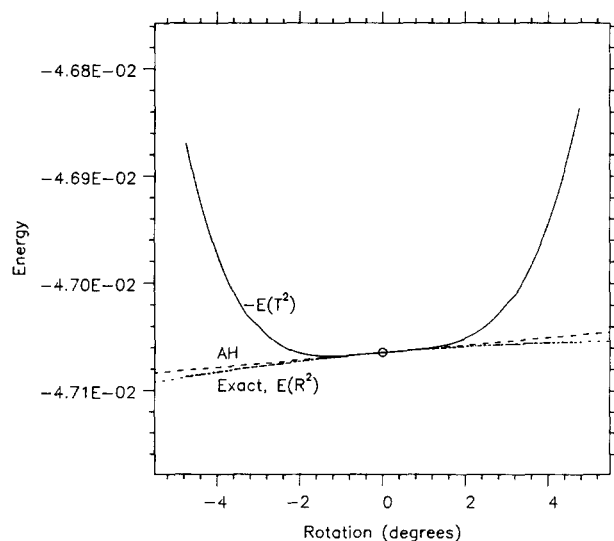


FIG. 2. The dependence of various energy approximations on the rotation $2\sigma-3\sigma$ for the HF molecules as in Fig. 1.

the undamped NR method [$E(R^2)$] predicts a step in the wrong direction. The undamped AH method predicts a step in the correct direction, which is, however, much too large. Of course, the step size can be reduced by adding a level shift, but this is completely arbitrary and therefore not shown. In contrast to the NR and AH approximations our second order energy functional $E(T^2)$ is in close agreement with the exact energy function. The approximation $E(T^2)$ is not only periodic as it should be, but also predicts the minima at nearly the correct angles. This illustrates the fast convergence of our method for internal-external rotations.

A very different situation is observed in Fig. 2 for a rotation between two strongly occupied orbitals. In this case the exact energy function is very flat, since the energy of the dominant (SCF) configuration is invariant with respect to this orbital rotation. Again, the NR and AH methods predict steps either in the wrong direction or too large ($+61^\circ$ and -542° , respectively; the optimum step is -53°). However, as shown in Fig. 2, the approximation $E(T^2)$ predicts only a very tiny step, even though for much larger angles $E(R^2)$ is still in close agreement with the exact energy. This illustrates the slow convergence of the method of Werner and Meyer for rotations between strongly occupied orbitals.

How can this effect be explained? Let us simply consider a two orbital model, e.g., the Be atom. Clearly, the energy of the Hartree-Fock determinant $|1s^2 2s^2|$ should be invariant with respect to any rotations between the orbitals $1s$ and $2s$. Defining a unitary transformation as $U = 1 + T$, with

$$T = \begin{pmatrix} X & S \\ -S & X \end{pmatrix}, \quad (38)$$

and expanding the two-electron energy

$$E_2^{(0)} = (11|11) + (22|22) + 4(11|22) - 2(12|12) \quad (39)$$

in powers of X and S , we obtain to fourth order in T ,

$$E_2^{(4)} = E_2^{(0)}[(1+X)^2 + S^2]^2. \quad (40)$$

Substituting $S = \sin \alpha$ and $X = \cos \alpha - 1$, one immediately finds that $E_2^{(4)}$ is invariant to changes in α ; however this is not true for our second-order approximation

$$E_2^{(2)} = E_2^{(0)}(1 + 4X + 6X^2 + 2S^2). \quad (41)$$

Approximating S and X to second order in α , one obtains

$$E_2^{(2)} = E_2^{(0)}(1 + \frac{3}{2}\alpha^4), \quad (42)$$

i.e., the functional behavior observed for $E(T^2)$ in Fig. 2. Due to the large positive prefactor this term strongly overcompensates the energy lowering due to the improvements of the less important configurations. This clearly explains why the $E(T^2)$ approximation is so bad for rotations between strongly occupied orbitals.

In the NR approximation, in which $E_2^{(2)}$ is approximated only to second order in α , the α^4 term vanishes, and, as also see in Fig. 2, the energy is invariant with respect to rotations between doubly occupied orbitals. This explains why the NR or AH methods exhibit better convergence than the method of Werner and Meyer once the starting approximation is close enough to the final solution.

However, as seen before, the AH method is often rather slowly convergent far from the solution. The only reasonable way to remedy the slow convergence of the method of Werner and Meyer appears therefore to be the inclusion of higher order terms in the energy approximation. In fact, to third order in S and X , the error in the two electron energy of the above Be model is the same as in second order, but in the opposite direction—an even worse situation. Hence, the internal-internal rotations have to be treated exactly to fourth order. As will be shown in the following, this can quite easily be accomplished.

First, we realize that for any unitary matrix U which transforms the internal orbitals only amongst themselves, the one electron Hamiltonian as well as the Coulomb and exchange operators can be transformed exactly as

$$h_{rs} \Rightarrow (U^\dagger h U)_{rs}, \quad (43)$$

$$(J^{kl})_{rs} \Rightarrow \sum_{ij} (U^\dagger J^{ij} U)_{rs} U_{ik} U_{jl}, \quad (44)$$

$$(K^{kl})_{rs} \Rightarrow \sum_{ij} (U^\dagger K^{ij} U)_{rs} U_{ik} U_{jl}. \quad (45)$$

This transformation requires about $\frac{3}{2}N^2M^4$ operations if performed in a brute-force way. Symmetry greatly reduces the effort required. On vector machines the transformation can be performed in terms of matrix multiplications (all $U_{ik}U_{jl}$ for a given k, l form a vector, all operators a supermatrix) and is therefore fast. If all operators of one symmetry type cannot be kept in high-speed memory simultaneously, the transformation can be performed in two or more passes; the number of I/O operations is proportional to the number of passes.

Hence we can start each macroiteration with an optimization of the internal-internal orbital rotations and CI coefficients only, using the step restricted Augmented

Hessian method. In each step, the integrals $(ij|kl)$ and h_{ij} , which are kept in core, are transformed (this is very cheap). After 1–3 orbital optimization steps, which converge quadratically, the CI coefficients are updated. This process, which can be done entirely in core (except for reading the formula tape) is repeated until the orbital and CI gradients are smaller than a certain threshold. Convergence of the CI coefficients is improved by the DIIS procedure.³⁶ At the end of the optimization one obtains a variational energy which has been minimized with respect to the internal–internal orbital rotations and variations of the CI coefficients. Using the final internal transformation matrix, the operators \mathbf{h} , \mathbf{J}^{kl} , and \mathbf{K}^{kl} are transformed according to equations (43)–(45). The orbitals are transformed correspondingly, and the calculation is then continued as described in Secs. II–IV.

The additional optimization of the internal orbital space in the beginning of each macroiteration improves convergence considerably. However, convergence can be improved even more by also minimizing the internal–internal orbital rotation parameters R_{ij} which are created by the coupling to the internal–external rotations. This can be accomplished by constructing an internal unitary matrix from the elements R_{ij} (i, j internal only), which are obtained after solving the nonlinear equations (18), and performing the transformations of Eqs. (43)–(45) again. [If \mathbf{U} is updated by successive application of Eq. (17), \mathbf{R} is not directly obtained. It can be well approximated, however, as $\mathbf{R} = \frac{1}{2}(\mathbf{U} - \mathbf{U}^\dagger)$.] Then, a few additional microiterations are required to solve the nonlinear equations (18) using the new operators. Of course, in this case a good starting approximation is available from the previous elements R_{ai} (a external only). This process can be repeated and usually converges rapidly (except sometimes in the first macroiteration, but then it is not important). Ideally, one ends up with a matrix \mathbf{R} in which all internal–internal elements are zero. This means that the internal–internal orbital rotations have been effectively treated to highest possible order with the operators \mathbf{J}^{kl} and \mathbf{K}^{kl} of the present macroiteration. Since the same is true for the optimization of the CI coefficients as described in the previous section, about the same convergence behavior is to be expected for CASSCF (no internal–internal orbital rotations) and more general MCSCF calculations. In practice, it is usually sufficient to absorb the internal–internal elements R_{ij} into the present orbital basis only once per macroiteration. The additional computational effort to perform the transformations (43)–(45) appear to be well justified, since the number of macroiterations, which require a much more expensive four-index transformation of the integrals, is considerably reduced. This will be demonstrated in the next section.

VI. NUMERICAL EXAMPLES

In this section we present some numerical examples which demonstrate that our new method exhibits considerably better convergence than all previous MCSCF optimization methods. Tables I–III show several simple test

TABLE I. Convergence of CASSCF calculations for the HF molecule (no internal–internal orbital rotations).^a

Iter.	Energy difference		
	AH ^b	WM ^c	New method
1	−0.013 386 636 0	−0.022 639 072 1	−0.029 157 802 9
2	−0.005 656 622 2	−0.005 729 368 4	−0.000 225 746 3
3	−0.008 010 545 8	−0.001 002 072 8	−0.000 000 000 5
4	−0.002 279 021 8	−0.000 012 721 3	
5	−0.000 050 679 2	−0.000 000 000 8	
6	−0.000 000 044 8		
7	−0.000 000 000 0		

^a Active space 2σ – 5σ (20 configurations), 1σ orbital frozen. Start with canonical Hartree–Fock orbitals. Basis set and geometry as for Fig. 1. Final energy: $-100.082\,638\,076\,9$ a.u.

^b Augmented Hessian method with step length control as described in Ref. 9. The steps are restricted to a norm of 0.5, 0.5, and 0.6 in the first three iterations, respectively.

^c Method of Werner and Meyer (Ref. 4).

calculations for the HF molecule, using canonical Hartree–Fock orbitals as starting guess. In Table I the augmented Hessian method with step length control as described by Jørgensen *et al.* (Fletcher optimization) is compared with the method of Werner and Meyer and our new procedure for a CASSCF calculation in which no internal–internal orbital rotations have to be performed (1σ orbital frozen). The AH calculation needs six iterations, the method of Werner and Meyer four iterations, and our new method only two iterations to converge the energy to within 10^{-9} a.u. The differences in convergence between the method of Werner and Meyer and our new method are in this case solely due to the different treatment of the coupling between the orbital and CI coefficient changes as described in Sec. IV. It is obvious that the higher order terms included in the present method considerably improve convergence.

Table II shows calculations for a wave function which comprises only the Hartree–Fock determinant and all double replacements from it into the 4σ and 5σ orbitals (11 configurations). In this case the rotations between the two strongly occupied orbitals 2σ , 3σ , and also between the two weakly occupied orbitals 4σ , 5σ are still redundant, but rotations between the strongly and weakly occupied orbitals have to be performed. Still, a similar fast convergence behavior as in Table I is observed for both the Werner–Meyer method and our new procedure.

A very different situation, however, is found in Table III. In this case the two configurations $1\sigma^2 2\sigma^2 4\sigma^1 5\sigma^1 1\pi^4$ and $1\sigma^2 3\sigma^2 4\sigma^1 5\sigma^1 1\pi^4$ have been omitted from the 11 configurations of Table II. The resulting 9 configuration MCSCF wave function gives the same final energy as the 11 configuration function of Table II, but in this case no internal orbital rotations are redundant. As expected from the findings of Sec. V, the method of Werner and Meyer converges slowly in this case, which is mainly due to the poor convergence of the 2σ – 3σ orbital rotation. Our new method, however, converges almost as fast as in the previous examples. This demonstrates the importance of the higher order treatment of the internal–internal orbital rotations. As in all following examples, the initial opti-

TABLE II. Convergence of test calculations for HF for a case with rotations between strongly and weakly occupied orbitals.^a

Iter.	Energy difference	
	WM ^b	New method ^c
1	-0.023 838 790 2	-0.029 156 144 8
2	-0.004 607 475 2	-0.000 241 229 8
3	-0.000 944 097 8	-0.000 000 000 7
4	-0.000 006 854 3	
5	-0.000 000 000 0	

^a All double excitations from SCF determinant into the 4σ and 5σ orbitals (11 configurations). For other details see footnote a of Table I (1σ not frozen in this case). Final energy: -100.082 533 343 5 a.u.

^b Method of Ref. 4.

^c One absorption of elements R_{ij} per iteration (see the text). Initial optimization of internal orbitals only in second iteration. Energy differences calculated from lowest variational energies of each iteration.

mization of the internal orbitals has been performed from only the second iteration; in the first iteration it is of little use, since the internal-internal orbital rotations strongly depend on the internal-external rotations. The "absorption" procedure described in the previous section, however, has been applied once in each macroiteration. We note that the energy change in the third iteration of this example can be reduced by a factor of 10 by performing more than one internal absorption.

Tables IV and V show comparisons of Fletcher optimizations recently reported by Jørgensen *et al.*³⁰ for the N₂ and CO molecules with our method. Configurations and basis sets are as in Ref. 30 (we were not able, however, to reproduce exactly the initial or final energies reported by Jørgensen *et al.* for CO). In Table IV the results for N₂ are presented; this calculation was considered by Jørgensen *et al.* as of "average complexity". So called IVO orbitals⁴¹ were used as starting guess (see footnote a of Table IV). The Fletcher optimization of Ref. 30 required six iterations, whereas our method converges in only three iterations. Table IV also shows the improvement in convergence on performing two internal absorptions

TABLE III. Convergence of test calculations for the HF molecule for a case with rotations between strongly occupied orbitals.^a

Iter.	Energy difference	
	WM ^b	New method ^c
1	-0.019 176 359 2	-0.030 658 214 3
2	-0.010 456 322 7	-0.000 339 908 2
3	-0.001 143 607 4	-0.000 000 006 5
4	-0.000 135 358 9	
5	-0.000 056 867 3	
6	-0.000 018 487 1	
7	-0.000 003 434 4	
8	-0.000 000 181 2	
9	-0.000 000 000 6	

^a Nine configurations (see the text). For other details see Table I. No orbitals frozen. Same final energy as in Table II.

^b Method of Ref. 4.

^c One absorption of elements R_{ij} per iteration (see the text). Initial optimization of internal orbitals only in second iteration. Energy differences calculated from lowest variational energies of each iteration.

TABLE IV. Convergence of test calculations for the N₂ molecule using IVO orbitals as starting guess.^a

Iter.	Energy difference		
	Fletcher ^b	New method ^c	New method ^d
1	-0.019 530 483 4	-0.052 436 691 2	-0.053 795 844 8
2	-0.016 995 391 4	-0.003 387 148 4	-0.002 049 568 4
3	-0.016 661 070 5	-0.000 023 448 7	-0.000 001 875 1
4	-0.002 740 679 3	-0.000 000 000 1	-0.000 000 000 0
5	-0.000 038 989 7		
6	-0.000 000 012 7		
7	-0.000 000 000 0		

^a Configurations and basis set as in Ref. 30. The IVO orbitals (Ref. 44) were obtained by diagonalizing, within σ , π_x , and π_y symmetries separately, the operator $h + \sum (2J_i - K_i) - J_i + 2K_i$, where i is the highest occupied orbital of the symmetry.

^b Fletcher optimization; results taken from Ref. 30. Final energy: -109.095 718 059 2 a.u.

^c See footnote c of Table III. Final energy: -109.095 718 128 8 a.u.

^d As footnote c, but two absorptions of elements R_{ij} per iteration.

in each macroiteration; however, it is not usually possible to save a macroiteration in this way, and so we do not recommend that more than one absorption be performed. Table V shows a similar calculation for the CO molecule, which was considered in Ref. 30 to be more difficult. The Fletcher optimization required 9–10 iterations, dependent on the initial maximum step size (only the faster calculation is shown in Table V). Our method turns out to be quite insensitive to the increased complexity, convergence being achieved as fast as in the previous examples. Even if canonical Hartree-Fock orbitals are used as starting guess (Table VI), which are much worse approximations than the IVO orbitals (because of the very diffuse functions in this basis set, the 6σ and 2π SCF orbitals are Rydberg

TABLE V. Convergence of test calculations for the CO molecule using IVO orbitals as starting guess.^a

Iter.	Energy difference	
	Fletcher ^b	New method ^c
1	-0.016 819 128 1	-0.056 168 287 7
2	-0.017 490 633 0	-0.002 605 967 7
3	-0.013 430 375 5	-0.000 001 306 1
4	-0.015 795 237 5	-0.000 000 000 0
5	-0.003 660 917 0	
6	-0.001 231 696 1	
7	-0.000 050 933 2	
8	-0.000 001 713 6	
9	-0.000 000 002 8	
10	-0.000 000 000 0	

^a 14 symmetry adapted configurations, see Ref. 30. [321] basis set of Ref. 44; one diffuse s and one p function added for each atom with exponents optimized for 3s, 3p in Ref. 44. The exponents of the d functions are 0.65 and 0.85 for C and O, respectively. Internuclear distance 2.132 bohr. IVO orbitals as for table IV, but using the operator $h + \sum j(2J_j - K_j) - J_i$.

^b Fletcher optimization; results taken from Ref. 30. Final energy: -112.880 980 331 9 a.u.

^c One absorption of elements R_{ij} per iteration, see text. Initial optimization of internal orbitals from second iteration. Energy differences calculated from lowest variational energies of each iteration. Final energy: -112.881 671 166 9 a.u.

TABLE VI. Convergence of test calculations for the CO molecule using canonical SCF orbitals as starting guess.^a

Iter.	AH ^b		New method ^d	
	Energy diff.	Step ^c	Energy diff.	Step ^c
1	-0.005 805 627 8	0.50	-0.114 267 389 6	2.3
2	-0.015 098 630 5	0.33	-0.006 158 512 1	0.36
3	-0.027 426 812 8	0.40	-0.000 002 589 2	0.62-2
4	-0.026 688 320 6	0.48	-0.000 000 000 0	0.12-4
5	-0.009 411 258 2	0.58		
6	-0.018 214 693 5	0.69		
7	-0.016 727 086 3	0.75		
8	-0.001 016 207 6	0.41		
9	-0.000 039 643 1	0.83-1		
10	-0.000 000 210 4	0.30-2		
11	-0.000 000 000 0	0.42-4		

^a Configurations and basis set as in Table V.^b Augmented Hessian calculation with step length control as described in Ref. 9. The steps are restricted in the first seven iterations.^c Step lengths are defined as $(\sum_{r,i} R_{ri}^2 + \sum_i \Delta c_i^2)^{1/2}$.^d See footnote c of Table V.^e Step lengths are defined as $(\sum_{r,i} R_{ri}^2)^{1/2}$, not including the absorbed internal rotations.

rather than antibonding in character), the energy converges to machine accuracy in three iterations. We note that the initial Hessian matrix has 21 negative diagonal elements, and probably a rather larger number of negative eigenvalues. Despite this, our method exhibits better than quadratic convergence right from the first iteration.

In Table VII we present a somewhat larger CASSCF calculation for formaldehyde. The wave function is comprised of 3644 configurations, and the basis set consists of 60 contracted functions. As before, the calculation converges in three iterations. Since in this case no internal-internal orbital rotations have to be performed (1s orbitals of C and O frozen), the calculation is particularly simple.

TABLE VII. Convergence of CASSCF calculations for formaldehyde.^a

Iter.	Energy difference		
	Without coupling ^b	With coupling ^c	With coupling ^d
1	-0.095 301 119 7	-0.106 133 147 2	-0.105 835 732 9
2	-0.009 714 001 7	-0.001 264 145 7	-0.001 561 529 3
3	-0.001 696 277 1	-0.000 000 013 6	-0.000 000 044 4
4	-0.000 446 948 3		
5	-0.000 141 130 9		
6	-0.000 052 972 8		
7	-0.000 022 910 5		
8	-0.000 010 824 3		
9	-0.000 005 361 1		

^a Active space $3a_1-7a_1, 1b_2-3b_2, 1b_1-2b_1$; 3644 configurations. Basis set: Huzinaga (Ref. 43) 9s, 5p (innermost 4s, 2p contracted) for C and O, augmented by single *d* functions with exponents 0.6 and 1.0, respectively; Huzinaga 5s (innermost 2s contracted) for H, augmented by one *p* function with exponent 0.75. Geometry: $R_{CH} = 1.116 \text{ \AA}$, $R_{CO} = 1.208 \text{ \AA}$, $\alpha_{HCH} = 109.471^\circ$. Final energy: $-114.033\,845\,498\,4 \text{ a.u.}$ ^b Equation (18) applied with fixed CI coefficients. The CI coefficients are optimized with $\mathbf{R} = \mathbf{0}$ at the beginning of each iteration.^c $E(T^2)$ fully optimized with respect to \mathbf{R} and Δc .^d Same as footnote c, but only five updates of CI coefficients in the first iteration.

In Table VII it is also shown that approximate second order schemes which omit the CI-orbital coupling may converge extremely slowly. It is found that the inclusion of the CI-orbital coupling considerably improves convergence right from the first iteration. However, in the first iteration it is sufficient to update the CI coefficients only a few times.

Finally, in Table VIII a simple calculation for LiH is shown, in which the average energy of the lowest two $^1\Sigma^+$ states has been optimized. Convergence is as fast as in the other examples. More elaborate calculations with very large numbers of configurations will be presented in a forthcoming paper.⁴²

VII. CONCLUSIONS

A new second order MCSCF method has been described which shows a remarkable improvement of convergence when compared with all previous methods. This is achieved by minimizing a second order energy approximation which takes into account the orthonormality of the orbitals exactly and is therefore periodic in the orbital rotations. Rotations between internal orbitals and the changes of the CI coefficients are treated effectively to highest possible order without performing a more expensive four index integral transformation than in any other second-order MCSCF method. The additional computational effort per iteration for the more accurate optimization of the internal orbital rotations is relatively small, and in most cases by far outweighed by the reduction in the number of macroiterations needed. It has been shown that the method can also be used to minimize an energy average of several states. Furthermore, since the Hessian and Hamiltonian matrices are never explicitly calculated and stored, very large basis sets and configuration expansions can be handled. Using a new method^{39,42} to reevaluate the coupling coefficients very efficiently each time they are needed, CASSCF calculations with up to about 10^5 configurations are possible.⁴² Since the whole formalism is based on matrix operations (mainly matrix multiplications), our method is perfectly well suited for the efficient use of modern vector computers.

For several examples it has been demonstrated that the method converges in 2-3 iterations to the required accuracy, even if poor starting orbitals are employed and the initial Hessian matrix has many negative eigenvalues.

TABLE VIII. Convergence of a state averaged MCSCF calculation for the LiH molecule.^a

Iter.	Energy difference ^b
1	-0.023 611 686 4
2	-0.000 040 358 8
3	-0.000 000 000 5

^a CASSCF calculation with active space $2\sigma-5\sigma$. The energy expression optimized was the equal weight average of the first two $^1\Sigma^+$ states. Basis set: [3s2p] contraction for Li, [2s] for H as described in Ref. 44. Geometry: $R = 3.0 \text{ a.u.}$ Final energies: $E_1 = -7.998\,214\,63 \text{ a.u.}$, $E_2 = -7.880\,722\,71 \text{ a.u.}$ ^b Successive changes in the average energy.

Since the observed convergence behaviour is better than quadratic right from the first iteration, it is unlikely that significantly better convergence can be obtained with any other second order optimisation procedure. Higher order methods, however, require an expensive full integral transformation in each iteration, and it is therefore unlikely that such methods would be more efficient. We therefore believe that in our method optimum convergence is achieved with a minimum of computational effort.

ACKNOWLEDGMENTS

HJW acknowledges grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie, which made possible this work. HJW also wishes to thank B. Lengsfeld for a useful discussion.

APPENDIX A: MULTISTATE DIRECT-CI TREATMENT

For cases in which some energetically low lying configurations are nearly degenerate, or if the energy average of several states is optimized, a simple partitioning technique can be used to update the CI coefficients. For this purpose, the configurations space $\{\Phi_I\}$ is divided into two subspaces $\{\Phi_P\}$ and $\{\Phi_Q\}$. The space $\{\Phi_P\}$ contains all configurations whose energies $\langle\Phi_P|H|\Phi_P\rangle$ are smaller than a prescribed threshold. The coefficients c_P of these "primary" configurations are determined conventionally by explicitly constructing and diagonalizing a Hamiltonian matrix, whereas the coefficients c_Q of the remaining "secondary" configurations are determined perturbationally in a direct-CI-like manner. The procedure can be performed as follows:

(1) Calculate

$$H_{Pn} = \sum_Q H_{PQ} c_Q^n, \quad (46)$$

$$H_{Qn} = \sum_{Q'} H_{QQ'} c_{Q'}^n, \quad (47)$$

$$H_{nm} = \sum_Q c_Q^n H_{Qm}, \quad (48)$$

and

$$S_{nm} = \sum_Q c_Q^n c_Q^m, \quad (49)$$

where the indices n, m denote all states required; hence, in a single state treatment H_{Pn} and H_{Qn} are vectors with the dimensions of the primary and secondary subspaces, respectively. The procedure is started with $c_Q^n = 0$.

(2) Solve the eigenvalue problem

$$\begin{pmatrix} H_{PP'} & H_{Pm'} \\ H_{mP'} & H_{mm'} \end{pmatrix} - E_n \begin{pmatrix} 1 & 0 \\ 0 & S_{mm'} \end{pmatrix} \begin{pmatrix} d^n \\ \alpha^n \end{pmatrix} = 0 \quad (50)$$

using standard techniques for all required states n .

(3) Form the vectors

$$g_Q^n = \sum_m \alpha_m^n H_{Qm} + \sum_P H_{QP} d_P^n, \quad (51)$$

$$d_Q^n = \sum_m \alpha_m^n c_Q^m. \quad (52)$$

(4) Update the CI coefficients as

$$c_P^n = d_P^n, \quad (53)$$

$$c_Q^n = d_Q^n - (g_Q^n - E_n d_Q^n)/(H_{QQ} - E_n). \quad (54)$$

(5) Orthonormalize the CI vectors and return to step (1).

This procedure requires the explicit calculation of the Hamiltonian matrix $H_{PP'}$ and two reads of the coupling coefficients γ_{ij}^{PQ} and Γ_{ijkl}^{PQ} in each microiteration. It is therefore recommended to sort the coupling coefficients (P, P) , (P, Q) and (Q, Q) onto three different files once the primary space has been selected. If several CI iterations are performed for a given U matrix, it is straightforward to extend the method to a Davidson-like procedure.³⁷

- ¹ E. Dalggaard and P. Jørgensen, *J. Chem. Phys.* **69**, 3833 (1978).
- ² D. L. Yeager and P. Jørgensen, *J. Chem. Phys.* **71**, 755 (1979).
- ³ B. H. Lengsfeld III, *J. Chem. Phys.* **73**, 382 (1980).
- ⁴ H.-J. Werner and W. Meyer, *J. Chem. Phys.* **73**, 2342 (1980).
- ⁵ H.-J. Werner and W. Meyer, *J. Chem. Phys.* **74**, 5794 (1981).
- ⁶ B. H. Lengsfeld III and B. Liu, *J. Chem. Phys.* **75**, 478 (1981).
- ⁷ R. Shepard, I. Shavitt, and J. Simons, *J. Chem. Phys.* **76**, 543 (1982).
- ⁸ B. H. Lengsfeld III, *J. Chem. Phys.* **77**, 4073 (1982).
- ⁹ H. J. Aa. Jensen, and P. Jørgensen, *J. Chem. Phys.* **80**, 1204 (1984).
- ¹⁰ K. Ruedenberg, L. M. Cheung, and S. T. Elbert, *Int. J. Quantum Chem.* **16**, 1069 (1979).
- ¹¹ B. Roos, P. Taylor, and P. E. M. Siegbahn, *Chem. Phys.* **48**, 157 (1980).
- ¹² P. E. M. Siegbahn, A. Heiberg, B. O. Roos, and B. Levy, *Phys. Scr.* **21**, 323 (1980).
- ¹³ P. E. M. Siegbahn, J. Almlöf, A. Heiberg, and B. O. Roos, *J. Chem. Phys.* **74**, 2384 (1981).
- ¹⁴ R. Shepard and J. Simons, *Int. J. Quantum Chem. Symp.* **14**, 211 (1980).
- ¹⁵ J. Olsen, D. L. Yeager, and P. Jørgensen, *Adv. Chem. Phys.* **54**, 1 (1983).
- ¹⁶ R. J. Buenker, S. D. Peyerimhoff, and W. Butscher, *Mol. Phys.* **35**, 771 (1978).
- ¹⁷ B. R. Brooks, W. D. Laidig, P. Saxe, N. C. Handy, and H. F. Schaefer III, *Phys. Scr.* **21**, 312 (1980).
- ¹⁸ P. E. M. Siegbahn, *J. Chem. Phys.* **72**, 1647 (1980).
- ¹⁹ B. Liu and M. Yoshimine, *J. Chem. Phys.* **74**, 612 (1981).
- ²⁰ H.-J. Werner and E. A. Reinsch, *J. Chem. Phys.* **76**, 3144 (1982).
- ²¹ H. Lischka, R. Shepard, F. B. Brown, and I. Shavitt, *Int. J. Quantum Chem. Symp.* **15**, 91 (1981).
- ²² V. R. Saunders and J. H. Van Lenthe, *Mol. Phys.* **48**, 923 (1983).
- ²³ D. J. Thouless, *The Quantum Mechanics of Many Body Systems* (Academic, New York, 1961).
- ²⁴ B. Levy, *Chem. Phys. Lett.* **4**, 17 (1969); *Int. J. Quantum Chem.* **4**, 297 (1970).

- ²⁵ D. L. Yeager and P. Jørgensen, *Mol. Phys.* **39**, 587 (1980).
- ²⁶ D. L. Yeager, P. Albertsen, and P. Jørgensen, *J. Chem. Phys.* **73**, 2811 (1980).
- ²⁷ P. Jørgensen, J. Olsen, and D. L. Yeager, *J. Chem. Phys.* **75**, 5802 (1981).
- ²⁸ J. Olsen, P. Jørgensen, and D. L. Yeager, *J. Chem. Phys.* **76**, 527 (1982).
- ²⁹ J. Olsen and P. Jørgensen, *J. Chem. Phys.* **77**, 6109 (1982).
- ³⁰ P. Jørgensen, P. Swanstrøm, and D. L. Yeager, *J. Chem. Phys.* **78**, 347 (1983).
- ³¹ J. T. Golab, D. L. Yeager, and P. Jørgensen, *Chem. Phys.* **78**, 175 (1983).
- ³² R. Fletcher, *Practical Methods of Optimization* (Wiley, New York, 1980), Vol. 1.
- ³³ L. G. Yaffe and W. A. Goddard III, *Phys. Rev. A* **13**, 1682 (1976).
- ³⁴ J. Olsen, P. Jørgensen, and D. L. Yeager, *J. Chem. Phys.* **77**, 356 (1982).
- ³⁵ M. R. Hoffmann, D. J. Fox, J. F. Gaw, Y. Osamura, Y. Yamaguchi, R. S. Grev, G. Fitzgerald, H. F. Schaefer III, P. J. Knowles, and N. C. Handy, *J. Chem. Phys.* **80**, 2660 (1984).
- ³⁶ P. Pulay, *Chem. Phys. Lett.* **73**, 393 (1980); *J. Comput. Chem.* **3**, 556 (1982).
- ³⁷ E. R. Davidson, *J. Comput. Phys.* **17**, 87 (1975).
- ³⁸ P. E. M. Siegbahn, *Chem. Phys. Lett.* **109**, 417 (1984).
- ³⁹ P. J. Knowles and N. C. Handy, *Chem. Phys. Lett.* **111**, 315 (1984).
- ⁴⁰ K. K. Docken and J. Hinze, *J. Chem. Phys.* **57**, 4928 (1972).
- ⁴¹ W. J. Hunt and W. A. Goddard III, *Chem. Phys. Lett.* **3**, 414 (1969).
- ⁴² P. J. Knowles and H.-J. Werner, *Chem. Phys. Lett.* (in press).
- ⁴³ S. Huzinaga, *J. Chem. Phys.* **42**, 1293 (1965).
- ⁴⁴ T. H. Dunning and P. J. Hay, in *Modern Theoretical Chemistry*, edited by H. F. Schaefer III (Plenum, New York, 1977), Vol. 3.