

A quadratically convergent multiconfiguration–selfconsistent field method with simultaneous optimization of orbitals and CI coefficients

HansJoachim Werner and Wilfried Meyer

Citation: *J. Chem. Phys.* **73**, 2342 (1980); doi: 10.1063/1.440384

View online: <http://dx.doi.org/10.1063/1.440384>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v73/i5>

Published by the AIP Publishing LLC.

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

A quadratically convergent multiconfiguration–self-consistent field method with simultaneous optimization of orbitals and CI coefficients

Hans-Joachim Werner

Institut für Physikalische und theoretische Chemie der Universität Frankfurt/M, West Germany

Wilfried Meyer

Institut für Physikalische Chemie der Universität Kaiserslautern, West Germany

(Received 29 December 1979; accepted 26 March 1980)

A quadratically convergent MC–SCF procedure is described which is based on the direct minimization of the energy. In comparison to the Newton–Raphson technique, which has previously been applied by several authors for orbital optimization, the convergence radius is much improved by taking into account in the energy expansion those parts of third and higher order terms which account exactly for the orthonormality constraints imposed on the orbitals. The nonlinear equations which define the improved orbitals are solved iteratively by a simple adaption of the Gauss–Seidel method. The coefficients of the configuration expansion can be optimized simultaneously with the orbitals, a necessary requirement for over-all quadratic convergence. The removal of redundant variables as well as useful approximations for the optimization of core orbitals are discussed. The convergence of the method is demonstrated to be much superior to classical Fock operator techniques and MC–SCF methods which are based on the generalized Brillouin theorem. The formalism is carried down to matrix operations and shows a simple structure.

I. INTRODUCTION

The Hartree–Fock self consistent field method has been proven to be a powerful tool for the calculation and understanding of many ground state properties of molecules in the vicinity of their equilibrium structure. However, in most cases the one determinant Hartree–Fock approximation is not adequate to describe properly the dissociation of molecular bonds. Also, many excited states cannot be represented by a single configuration wave function. In order to calculate properties of such states or to investigate the formation of molecular bonds one needs multiconfiguration wave functions for which both, the linear coefficients c_I of the configuration expansion (in the following referred to as “CI coefficients”)

$$\Psi = \sum_I c_I \Phi_I \quad (1)$$

as well as the set of orthonormal molecular orbitals $\{\phi_i\}$, from which the configurations Φ_I are constructed, are optimized according to the variational principle. As well known, this “MC–SCF” problem presents much more difficulties than the simple one determinant Hartree–Fock case and much work has been devoted to it during the last decade.

The difficulties mainly arise from the fact that for general MC–SCF wave functions the energy is not invariant with respect to rotations between occupied orbitals. Hence, instead of a relatively simple pseudo eigenvalue equation in the one determinant case, the set of coupled Fock equations

$$\sum_j F_{ij} |\phi_j\rangle = \sum_j \epsilon_{ij} |\phi_j\rangle, \quad (2)$$

with the hermiticity conditions

$$\epsilon_{ij} = \epsilon_{ji}^* \quad (3)$$

has to be solved. ϵ_{ij} are Lagrange multipliers which account for the orthonormality constraints imposed on the orbitals. The Fock operators F_{ij} depend on the orbitals $\{\phi_i\}$ and the set of CI coefficients $\{c_I\}$. In analogy to the one determinant case many attempts have been made to solve these equations iteratively by keeping the Fock operators fixed in each interaction step.^{1–18} Then the Lagrange multipliers can be expressed by coupling operators⁷ constructed such that the Fock equations are transformed into pseudo eigenvalue equations yielding the improved orbitals. These are used in a second step to determine new CI coefficients by diagonalizing the CI matrix $\langle \Phi_I | \hat{H} | \Phi_J \rangle$. The convergence of these algorithms was mostly found to be poor. As pointed out by Hinze,¹⁷ this is partly due to the fact that the coupling operators lead to equations of high degree in the unknown variables. Furthermore, as already shown in 1969 by Huzinaga⁷ and more recently discussed by Hirao and Nakatsuji¹¹ the coupling operators cannot be defined uniquely. They depend on some unknown parameters which can crucially influence convergence.^{11,14,15} As will be discussed in more detail in Sec. V the neglect of the change of the Fock operators as functions of the orbitals and CI coefficients is a third source for poor convergence, in particular if rotations between occupied orbitals are performed. This is also true for the method recently devised by Hinze and Yurtsever¹⁸ which consists of solving Eq. (2) by a Jacobi-like rotation technique and thus avoids the problems arising from the coupling operators. An improved MC–SCF procedure has been developed by Das and Wahl.^{9,10,19} In their method the coupling operator technique is used only to improve the wave function with respect to rotations into the virtual orbital space. For rotations between occupied orbitals, however, the change of the Fock operators is approximately taken into account by expanding the Fock equations up to first order

in corresponding orbital rotation parameters, which can be determined by solving the resulting set of linear equations. This method has been applied successfully to quite a large number of problems,¹⁹ but to our knowledge its convergence behavior has never been demonstrated. The convergence radius seems to be limited since a tedious procedure to calculate suitable start orbitals has to be applied in advance to the MC-SCF calculation.⁹

A second group of MC-SCF methods²⁰⁻²³ is based on the generalized Brillouin theorem.²⁴ In these methods the orbital changes are derived from the coefficients of a "super CI" consisting of the MC-SCF wave function and all Brillouin singly excited configurations. These methods seem to converge quite well although not quadratically as sometimes stated (see Sec. V). They suffer, however, from the fact that large Hamilton and overlap matrices have to be calculated and diagonalized in each iteration step. The information needed to construct the Hamilton matrix is provided by the density matrices of the MC-SCF function up to fourth order (see Appendix C) which shows that this step is by no means trivial.

More recently, various methods²⁵⁻³⁴ have been proposed which are based on the direct minimization of the energy. In the methods of Yaffe and Goddard,³⁰ Kendrick and Hillier³² and Dalgaard, Yeager and Jørgensen^{33,34} the energy is expanded up to second order in some orbital rotation parameters from which the unitary transformation matrix for the orbitals can be constructed. Variation of the quadratic energy approximation with respect to the parameters leads to a system of linear equations for the rotation parameters which define the set of improved orbitals. This "Newton-Raphson" orbital optimization method is quadratically convergent. However, its convergence radius appears to be rather limited since the (Hessian) matrix of second energy derivatives must be positive definite. This is often not the case if one starts with Hartree-Fock orbitals.

In most methods described so far the orbitals and CI coefficients are optimized alternately: for a given set of molecular orbitals the CI coefficients are obtained by solving the Schrödinger Equation in configuration space. Then the coefficients are held fixed and the orbitals are further improved. This alternation destroys the over-all quadratic convergence, even if quadratically convergent orbital optimization techniques are applied. Actually, we found that under certain circumstances the convergence of the alternate process is very slow, for instance in cases where the wave function contains configurations which differ by single excitations only. In order to improve convergence in such cases the orbitals and CI coefficients must be optimized simultaneously. Such procedures have been discussed by Das,¹⁰ Chang and Schwarz²² and Dalgaard and Jørgensen³³ in the context of convergence difficulties for excited states. It should be emphasized, though, that the simultaneous optimization is necessary to achieve over-all quadratic convergence and is therefore often advantageous also in ground state calculations. This point has also been stressed in the most recent paper by Yeager and Jørgensen³⁴ which appeared during the preparation of this manuscript. Al-

though quadratic convergence behavior often begins only relatively closely to the final solution, it is important for mainly three reasons: first, the step requiring most computation time in any MC-SCF method is the (partial) transformation of the two electron integrals into MO basis which is required in each iteration. As compared to Fock operator methods the integral transformation required to achieve quadratic convergence is about twice as expensive. This additional effort is usually overcompensated by the drastic reduction of iterations. It should be noted that the "generalized Brillouin theorem MC-SCF methods" require the same integral transformation although they are not quadratically convergent. Second, for calculating properties the wave function should be well converged; third, when calculating energy surfaces the solution of a neighboring point usually provides very good starting orbitals. Quadratic convergence behavior is then immediately reached and thus improves efficiency considerably.

In this paper we present a general quadratically convergent MC-SCF orbital optimization procedure which is also based on the direct minimization of the energy. In contrast to the Newton-Raphson method as applied by Yaffe and Goddard,³⁰ Kendrick and Hillier³² and Yeager and Jørgensen³⁴ higher order terms, which account exactly for the orthonormality constraints, are taken into account in the energy expansion. In this way the true energy surface, which is a periodic function in the orbital rotations, is actually approximated by a periodic function. A minimum of this approximation can even be found if the Hessian matrix has negative eigenvalues at the point of expansion. This considerably improves the convergence radius as compared to the Newton-Raphson method without requiring any noteworthy additional computational effort. The nonlinear equations are solved by an adaption of the static Gauss-Seidel method³⁶ so that expensive matrix inversions are avoided. The method becomes identical to the Newton-Raphson technique if all higher order terms are left out and becomes equivalent to Hinze's Fock operator method¹⁸ if the changes of the Fock operators are neglected. As an important option the orbitals and CI coefficients can be optimized simultaneously. It is shown that this requires only little additional computational effort but drastically improves convergence if the alternate process is slow. Since all second energy derivatives with respect to the orbital and the CI coefficients are taken into account exactly, the procedure is quadratically convergent for general MC-SCF wave functions.

In Sec. II we present our orbital optimization technique for fixed CI coefficients. The simultaneous optimization of the orbitals and the coefficients is explained in Sec. III. In Sec. IV we present an approximate treatment of energetically low lying core orbitals which are doubly occupied in all determinants. This considerably reduces the computational effort without much influencing the fast convergence of the method. In Sec. V we compare our method to previous work and some examples for the convergence behavior of various methods will be given.

II. THEORY OF QUADRATICALLY CONVERGENT ORBITAL OPTIMIZATION

In this Section we will present our formalism of quadratically convergent MC-SCF orbital optimization for fixed coefficients of the configuration expansion (1). The molecular orbitals (MO's) are assumed to be real linear combinations of a set of N real basis functions (AO's) χ_μ :

$$\phi_i = \sum_{\mu} \chi_{\mu} C_{\mu i} . \quad (4)$$

We will denote all orbitals which are occupied in any of the determinants "internal" or "occupied" orbitals, all others "external" or "virtual" orbitals. For a given basis set $\{\chi\}$ the orbital ϕ_i is represented by the i th column of the coefficient matrix C . In the following this column vector will be represented by the Dirac "ket" $|i\rangle$, whereas the corresponding row vector is denoted by the "bra" $\langle i|$. Thus, for any operator \hat{O} , the integral $\int \phi_i \hat{O} \phi_j d\tau$ is given by

$$\langle i | \hat{O} | j \rangle = \sum_{\mu\nu} C_{\mu i} O_{\mu\nu} C_{\nu j} = (C^* \hat{O} C)_{ij} , \quad (5)$$

where the matrix O has the elements

$$O_{\mu\nu} = \int \chi_{\mu} \hat{O} \chi_{\nu} d\tau . \quad (6)$$

The orthonormality condition imposed on the molecular orbitals takes the form

$$\int \phi_i \phi_j d\tau = \langle i | S | j \rangle = \delta_{ij} \quad (7)$$

where $S_{\mu\nu} = \int \chi_{\mu} \chi_{\nu} d\tau$ is the overlap matrix of the basis set $\{\chi\}$. The wave function is assumed to be spin restricted in the sense that the same set of orbitals is used for α and β spin orbitals. For a given set of MO's the total energy of the wave function (1) can generally be written in the form

$$E_0 = \sum_{ij} \langle i | h | j \rangle \gamma_{ij} + \frac{1}{2} \sum_{kl} \Gamma_{ij,kl} J_{kl} | j \rangle . \quad (8)$$

h is the one electron hamilton matrix; J_{kl} are generalized Coulomb operators

$$(J_{kl})_{\mu\nu} = \sum_{\rho\sigma} C_{\rho k} C_{\sigma l} (\mu\nu | \rho\sigma) , \quad (9)$$

with

$$(\mu\nu | \rho\sigma) = \int \chi_{\mu}(1) \chi_{\nu}(2) \frac{1}{r_{12}} \chi_{\rho}(1) \chi_{\sigma}(2) d1 d2 . \quad (10)$$

For later use we also define exchange operators

$$(K_{kl})_{\mu\nu} = \sum_{\rho\sigma} C_{\rho k} C_{\sigma l} (\mu\rho | \sigma\nu) . \quad (11)$$

Note the relations

$$\langle i | J_{kl} | j \rangle = (ij | kl) = (ji | kl) = (kl | ij) \quad (12)$$

$$\langle i | K_{kl} | j \rangle = (ik | lj) \quad (13)$$

and, hence

$$J_{kl} = J_{lk} , \quad K_{kl} = K_{lk}^* . \quad (14)$$

γ_{ij} and $\Gamma_{ij,kl}$ are elements of the spinfree first and sec-

ond order reduced density matrices, respectively. Due to the symmetry (12) of the orbital integrals Γ may be defined symmetric in the exchange of pairs of indices (compare Appendix C):

$$\gamma_{ij} = \gamma_{ji} \quad (15)$$

$$\Gamma_{ij,kl} = \Gamma_{ji,kl} = \Gamma_{kl,ij} . \quad (16)$$

The density matrices depend on the coefficients of the configuration expansion:

$$\gamma_{ij} = \sum_I C_I C_J \gamma_{ij}^{IJ} \quad (17)$$

$$\Gamma_{ij,kl} = \sum_{IJ} C_I C_J \Gamma_{ij,kl}^{IJ} , \quad (18)$$

with the normalization condition $\sum_I C_I^2 = 1$. The quantities γ_{ij}^{IJ} and $\Gamma_{ij,kl}^{IJ}$ are defined by

$$\begin{aligned} \langle \Phi_I | \hat{H} | \Phi_J \rangle &= \sum_{ij} \left\{ \gamma_{ij}^{IJ} \langle i | h | j \rangle \right. \\ &\quad \left. + \frac{1}{2} \sum_{kl} \Gamma_{ij,kl}^{IJ} \langle i | J_{kl} | j \rangle \right\} . \end{aligned} \quad (19)$$

They depend on the configurations $\{\Phi_I\}$ of the wave function and can be calculated using algorithms familiar in CI techniques. Note that γ_{ij} or $\Gamma_{ij,kl}$ vanish if any index corresponds to an external orbital.

According to the variational principle we now look for a unitary transformation U (more precisely an orthogonal transformation since we use only real orbitals) of the orbitals ϕ_i such that the transformed orbitals ϕ'_i minimize the energy expectation value (8):

$$|i'\rangle = \sum_k |k\rangle U_{ki} \quad (20)$$

or

$$C' = CU . \quad (21)$$

U can be expressed as

$$U = e^R = 1 + R + \frac{1}{2} R^2 + \dots , \quad (22)$$

where the antisymmetric matrix $R = -R^*$ represents a set of independent orbital rotation parameters for which we take the elements R_{ij} with $i > j$. Since the energy expression is a fourth order function of U it is in practice not possible to find U in one step. In earlier treatments³⁰⁻³⁴ the energy has therefore been expanded up to second order in R . Variation of this energy approximation with respect to the parameters R_{ij} leads to a system of linear (Newton-Raphson) equations from which the parameters can be calculated. This procedure, which has to be iterated, is only convergent if the matrix of second energy derivatives with respect to the R_{ij} is positive definite. Since the true energy surface is a periodic function in the orbital rotations, this is only the case if one starts relatively close to the final result. In order to improve the convergence radius we expand the energy in the orbital changes

$$|\Delta i\rangle = |i'\rangle - |i\rangle = \sum_k |k\rangle T_{ki} \quad (23)$$

which are described by the matrix

$$\mathbf{T} = \mathbf{U} - 1. \quad (24)$$

Using the relations (12)–(16) it is straightforward to derive from (8) the energy expression to second order in \mathbf{T} :

$$E^{(2)} = E_0 + 2 \sum_{ijk} T_{ki} \langle k | \mathbf{F}_{ij} | j \rangle + \sum_{ijkl} T_{ki} \langle k | \mathbf{G}_{ij} | l \rangle T_{li}. \quad (25)$$

The operators \mathbf{F}_{ij} and \mathbf{G}_{ij} are defined as follows:

$$\mathbf{F}_{ij} = \mathbf{h}_{ij} + \sum_{kl} \Gamma_{ij,kl} \mathbf{J}_{kl} \quad (26)$$

$$\mathbf{G}_{ij} = \mathbf{F}_{ij} + 2 \sum_{kl} \Gamma_{ik,jl} \mathbf{K}_{kl}. \quad (27)$$

The definition of the operators \mathbf{F}_{ij} is the same as in Hinze's treatments.^{17,18} Note the relations $\mathbf{F}_{ij} = \mathbf{F}_{ji}$ and $\mathbf{G}_{ij} = \mathbf{G}_{ji}^*$. The operators vanish if not both indices i and j correspond to internal orbitals. They can all be calculated from the set of "internal" operators \mathbf{J}_{ij} and \mathbf{K}_{ij} (i, j corresponding to internal orbitals). Thus, a complete transformation of the two electron integrals into orbital basis is not required. An efficient algorithm for the calculation of the operators \mathbf{J}_{ij} and \mathbf{K}_{ij} is described in Appendix A.

At the point of expansion $E^{(2)}$ has the same first and second derivatives with respect to the R_{ij} as the true energy surface. In contrast to the Newton–Raphson approximation the above energy expression is periodic in \mathbf{R} and a minimum can therefore always be found. For the minimum of $E^{(2)}$ with respect to \mathbf{T} the following equations must be satisfied:

$$\frac{\partial}{\partial T_{nm}} \left(E^{(2)} - \sum_{ij} \epsilon_{ij} \{ (\mathbf{U}^* \mathbf{U})_{ij} - \delta_{ij} \} \right) = 0 \text{ for all } n, m \quad (28)$$

with

$$\epsilon_{ij} = \epsilon_{ji}. \quad (29)$$

The terms involving the Lagrange multipliers ϵ_{ij} account for the orthonormality constraints (7). The restriction (29) follows from the symmetry $(\mathbf{U}^* \mathbf{U})_{ij} = (\mathbf{U}^* \mathbf{U})_{ji}$. From Eqs. (28) and (29) it is straightforward to derive the equations

$$2 \sum_j \langle n | \mathbf{F}_{mj} | j \rangle + 2 \sum_{jk} \langle n | \mathbf{G}_{mj} | k \rangle T_{kj} - 2(\mathbf{U} \epsilon)_{nm} = 0 \text{ for all } n, m. \quad (30)$$

Defining the matrices

$$A_{nm} = 2 \sum_j \langle n | \mathbf{F}_{mj} | j \rangle \quad (31)$$

$$B_{nm} = A_{nm} + 2 \sum_{jk} \langle n | \mathbf{G}_{mj} | k \rangle T_{kj}. \quad (32)$$

Equation (30) can be written in matrix form

$$\mathbf{B} - 2\mathbf{U}\epsilon = 0. \quad (33)$$

Note that $A_{nm} = 0$ and $B_{nm} = 0$ if m corresponds to an external orbital. Using the relations (29) and $\mathbf{U}^* \mathbf{U} = 1$ we

finally obtain the simple stationary condition for $E^{(2)}$:

$$\mathbf{E}' = \mathbf{U}^* \mathbf{B} - \mathbf{B}^* \mathbf{U} = 0. \quad (34)$$

Note that \mathbf{B} depends on $\mathbf{T} = \mathbf{U} - 1$. The matrix \mathbf{U} obeying (34) defines a new set of orbitals according to Eq. (20). The expansion $E \rightarrow E^{(2)}$ and minimization of $E^{(2)}$ is iterated until all orbital changes are smaller than a prescribed threshold. After convergence of the orbital optimization Eq. (34) must be valid for $\mathbf{R} = \mathbf{T} = 0$. This immediately yields the necessary conditions to be satisfied by the optimum orbitals:

$$\left(\frac{\partial E}{\partial R_{ij}} \right)_{\mathbf{R}=0} = (\mathbf{A} - \mathbf{A}^*)_{ij} = 0 \quad (35)$$

for all independent variables R_{ij} . Using the definition of the matrix \mathbf{A} , Eq. (35) can be written in the more familiar form

$$\sum_k (\langle i | \mathbf{F}_{jk} | k \rangle - \langle j | \mathbf{F}_{ik} | k \rangle) = 0 \text{ for all } i, j. \quad (36)$$

At this point we may discuss the problem of redundant variables which has been addressed by several authors.^{23,28,33} If the energy is invariant with respect to the rotation R_{ij} , e.g., for the case that the orbitals ϕ_i , ϕ_j are doubly occupied in all determinants, this rotation parameter is redundant. More generally, if the rotation between the orbitals ϕ_i and ϕ_j can be accounted for simply by changing the coefficients c_I , i.e., if

$$\frac{\partial \Psi(c_I)}{\partial R_{ij}} = \sum_I c_I^I \Phi_I, \quad (37)$$

the parameter R_{ij} is also redundant. Since Eq. (35) must be fulfilled for such variables as long as the c_I in Eq. (37) represent an eigenvector of the MC–CI, a value of $(\mathbf{A} - \mathbf{A}^*)_{ij}$ below a certain small threshold (e.g. 10^{-9}) indicates a redundant R_{ij} . Thus, an explicit search of hidden redundancies, which may be very complicated, is not required. The redundancies may simply be removed by keeping the corresponding R_{ij} equal to zero. It should be noted that convergence may be strongly impaired if redundancies are not considered. Obviously, Eq. (35) is also satisfied if ϕ_i and ϕ_j belong to different symmetries or are both external orbitals.

In the above derivation of the stationary condition (34) the orthonormality constraints (7) have formally been accounted for by introducing the Lagrange multipliers ϵ_{ij} . This treatment has the advantage of being very simple and reveals, as we will show in Sec. V, directly the connection of our method to the Fock operator techniques. Using the ansatz (22) for the unitary matrix \mathbf{U} it is also possible to derive an equivalent minimization condition without the use of the Lagrange multipliers. Inserting $\mathbf{T} = \mathbf{R} + \frac{1}{2}\mathbf{R}^2 + \dots$ into Eq. (25) and deriving with respect to R_{ij} one obtains the stationary condition

$$\mathbf{E}' = \mathbf{B}_- - \frac{1}{2}(\mathbf{B}_+ \mathbf{R} + \mathbf{R} \mathbf{B}_+) + \frac{1}{8}(\mathbf{B}_- \mathbf{R}^2 + \mathbf{R} \mathbf{B}_- \mathbf{R} + \mathbf{R}^2 \mathbf{B}_-) - \dots = 0 \quad (38)$$

where

$$E'_{ij} = \frac{\partial E^{(2)}}{\partial R_{ij}} \quad (39)$$

and

$$\mathbf{B}_\pm = \mathbf{B} \pm \mathbf{B}^* . \quad (40)$$

One can show, that the solutions of Eqs. (34) and (38) are the same. If \mathbf{R} is small one may neglect all second and higher order terms in Eq. (28). This leads to the linear Newton-Raphson equations

$$\mathbf{E}' = \mathbf{B}' - \frac{1}{2}(\mathbf{A}_\pm \mathbf{R} + \mathbf{R} \mathbf{A}_\pm) = 0 \quad (41)$$

with

$$\mathbf{B}'_{nm} = \mathbf{A}_{nm} + 2 \sum_{jk} \langle n | \mathbf{G}_{mj} | k \rangle \mathbf{R}_{kj} . \quad (42)$$

These equations are equivalent to those applied by Jaffe and Goddard³⁰ and Jaeger and Jørgensen.³⁴ Slightly different equations have been used by Kendrick and Hillier³² (note that the second energy derivative matrix as given in Ref. 32 is not symmetric and correct only for the final solution). Due to the definition of the operators \mathbf{F}_{ij} and \mathbf{G}_{ij} and the matrices \mathbf{A} and \mathbf{B} , Eqs. (41) and (42) appear to be of simpler structure than the corresponding ones in Refs. 30 and 32. In particular, there are no 4-index tensors involved and no programming or storage difficulties as mentioned in Ref. 32 occur. Jørgensen *et al.*^{33,34} have formulated Eq. (41) in the language of second quantization. The energy derivatives are then obtained in terms of Hamilton matrix elements between the MC-reference function and singly and doubly substituted configurations. This formulation leads to unnecessary complications and seems not to be very practical (see Sec. V and Appendix C).

As already stated, the necessary conditions for convergence of the Newton-Raphson method is that the matrix of second energy derivatives (τ_{ij} permutes the indices i and j , $i > j$, $k > l$)

$$H_{ij,kl} = \left(\frac{\partial^2 E}{\partial R_{ij} \partial R_{kl}} \right)_{R=0} = (1 - \tau_{ij})(1 - \tau_{kl}) \times \{ 2 \langle i | \mathbf{G}_{jl} | k \rangle - \frac{1}{2} A_{+ik} \delta_{jil} \} \quad (43)$$

is positive definite. This is frequently not the case if one starts with Hartree-Fock orbitals. Even if the Hessian matrix is positive definite divergence can be encountered if one starts near a point of inflection on the energy surface. Some second energy derivatives $H_{ij,ij}$ are then very small and the solution of Eq. (41) leads to much too large parameters R_{ij} . Actually, since the true energy surface is a periodic function of R_{ij} , the Newton-Raphson method, which assumes the second derivatives to be constant, generally tends to overestimate the orbital rotations. Of course, one can always achieve convergence by keeping the R_{ij} sufficiently small and of correct sign, but such trial and error procedure is often time consuming and a source of frequent annoyance. Such convergence difficulties do not arise for the present method which is based on a periodic energy approximation. Although far from the desired solution there is no *a priori* reason that our method must converge, we never encountered a divergent situation so far (usually starting with Hartree-Fock orbitals). The method was found to work well also in cases where the matrix of second energy derivatives had up to 40 negative eigenvalues. Convergence is usually very fast (1-4 iterations)

if only intern-extern rotations have to be performed. For intern-intern rotations, however, the optimization sometimes proceeds more slowly and the convergence behavior becomes quadratic only relative close to the final results. Near the final solution, where the Hessian matrix is positive definite, we found it then advantageous to apply Eq. (41) rather than Eq. (34). This effect must be attributed to the incomplete inclusion of the third and higher order terms with respect to the parameters R_{ij} in the expansion (25). This may result in a too large periodicity of $E^{(2)}$.

Eq. (34) or (41) may be solved iteratively by assuming that the diagonal second energy derivatives

$$D_{ij}^0 = H_{ij,ij} \quad (44)$$

are large as compared to the off-diagonal derivatives $H_{ij,kl}$ and do not depend on \mathbf{R} . This yields the iterative sequence:

$$R_{ij}^{(1)} = -(\mathbf{A} - \mathbf{A}^*)_{ij} / D_{ij}^0 \quad (45)$$

$$\Delta R_{ij}^{(n)} = -E'_{ij}^{(n)} / D_{ij}^0 \quad (46)$$

$$\mathbf{R}^{(n+1)} = \mathbf{R}^{(n)} + \Delta \mathbf{R}^{(n)} . \quad (47)$$

For the case that \mathbf{E}' is linear in \mathbf{R} [Eq. (41)] this is the static Gauss-Seidel method.³⁶ Unfortunately, some off-diagonal second energy derivatives (usually between intern-intern rotations) are often quite large. Furthermore, far from convergence some second derivatives D_{ij}^0 can be small or negative indicating that one starts near an inflection or saddle point on the energy surface. In such cases the above process is not or only very slowly convergent. We found, however, the following procedure to be satisfactory convergent:

(a) Instead of keeping D_{ij}^0 fixed, approximate $D_{ij}^{(n)}$ by the difference quotient

$$D_{ij}^{(n)} = (E'_{ij}^{(n)} - E'_{ij}^{(n-1)}) / (R_{ij}^{(n)} - R_{ij}^{(n-1)}) . \quad (48)$$

Take care, however, that $D_{ij}^{(n)}$ is always positive and does not change too fast, e. g.,

$$\frac{1}{3} D_{ij}^{(n-1)} < D_{ij}^{(n)} < 3 D_{ij}^{(n-1)} . \quad (49)$$

(b) In order to find reasonable start values $R_{ij}^{(1)}$ and $D_{ij}^{(1)}$ if D_{ij}^0 is small or negative, assume the energy to be a periodic function in $R_{ij}^{(1)}$, i. e.,

$$E = E_m \sin(n R_{ij}^{(1)}) \quad 1 \leq n \leq 4 . \quad (50)$$

It is then straightforward to determine $R_{ij}^{(1)}$ and $D_{ij}^{(1)}$ from the first and diagonal second energy derivatives. A value of $n=4$ was found to be appropriate in most cases.

(c) The parameter changes $\Delta R_{ij}^{(n)}$ should be kept smaller than a value of about 0.1 in each iteration step. If the linear equations (41) are used $R_{ij}^{(n)}$ should not be increased above a certain threshold.

(d) The linear Eq. (41) can in principle be solved in one step by inverting the Hessian matrix $H_{ij,kl}$. Unfortunately, this matrix is often very large and cannot be stored completely in high speed memory. Its inversion is also more expensive than the above iterative algorithm. However, the convergence of the latter can be improved by inverting only that part of the Hessian matrix which corresponds to rotations between internal

and possibly a few external orbitals (see Appendix B). We note that the matrix of second energy derivatives can be constructed during one sequential read of the operators G_{ij} . The rows and columns of the Hessian matrix which correspond to redundant variables R_{ij} should be eliminated, i.e., set equal to zero.

We may now summarize the MC-SCF orbital optimization process as follows:

- (1) Specify configurations.
- (2) Calculate the density matrix derivatives γ_{ij}^{IJ} and $\Gamma_{ij,kl}^{IJ}$ (only nonzero elements are stored together with their indices $i \leq j$, $k \leq l$, $(ij) \leq (kl)$).
- (3) For the given set of orbitals calculate the operators J_{ij} and K_{ij} and transform them into orbital basis. Transform the one electron Hamilton matrix h into orbital basis.
- (4) Using the density matrices γ^{IJ} and Γ^{IJ} and the operators J_{kl} , calculate according to Eq. (19) the Hamilton matrix $\langle \Phi_I | \hat{H} | \Phi_J \rangle$. Diagonalize it in order to find the CI coefficients c_I of the desired root.
- (5) Evaluate the density matrices γ and Γ according to Eqs. (17) and (18). Note that the number of independent nonzero elements $\Gamma_{ij,kl}$ is usually not very large and can be kept in high speed memory without difficulty.
- (6) Construct the matrices F_{ij} from the operators h and J_{kl} as calculated in step 3. In order to reduce I/O time as many matrices J_{kl} (of one particular symmetry type) as possible are first read into high speed memory. Then their contributions are added to the matrices F_{ij} which are treated one after another. This procedure is repeated until all J_{kl} have been read (in most cases all matrices J_{kl} can be kept in core simultaneously). Using the matrices F_{ij} and K_{kl} the matrices G_{ij} are then calculated and stored. A similar procedure as described above can be applied.
- (7) Solve Eq. (34) or (41) in order to obtain U . In each step of the iterative Gauss-Seidel type algorithm described above, the matrices $U^{(n)}$ and $B^{(n)}$ must be evaluated. $U^{(n)}$ can be approximated by truncating the power expansion (22) after the R^2 or R^3 term. The construction of B requires one sequential read of the matrices G_{ij} . In order to reduce I/O time one should keep as many matrices G_{ij} as possible in high speed memory during the iteration.
- (8) Transform the orbital coefficient matrix C according to Eq. (21). Since U is not exactly unitary apply the Schmidt orthonormalization procedure to the new coefficient matrix C' .
- (9) Return to step 3.

Direct access devices can be used to store all matrices. We note that except for the calculation of the operators J_{ij} and K_{ij} , the computational effort of all steps is only proportional to $n^2 N^2$ where n is the number of occupied orbitals and N is the number of basis functions. The evaluation of the operators J_{ij} and K_{ij} and their transformation into orbital basis requires about $3/8 N^4 n + 35/12 N^3 n^2 + 11/3 N^3 n$ multiplications (see Appendix A) and

is therefore for large basis sets the rate determining step. For comparison we note that the calculation of the closed shell Hartree-Fock operator requires about $3/4 N^4$ multiplications. Thus, assuming $N \cong 3n$ the MC-SCF calculation can be expected to take about $2n$ times more CPU time per iteration than a simple Hartree-Fock calculation. The number of iterations is, however, usually much smaller than for a Hartree-Fock calculation, since the MC-SCF procedure is quadratically convergent. The CPU time can be reduced by an approximate treatment of orbitals which are doubly occupied in all determinants (see Sec. IV). Note that the computation time is predominantly determined by the number of occupied (valence) orbitals but not by the number of configurations. The latter only influences the number of nonzero density matrix elements and, hence, the small effort for the calculation of the operators F_{ij} and G_{ij} . Although the Gauss-Seidel algorithm applied for the solution of Eq. (34) or (41) typically requires 10-20 iterations it is usually fast compared to the integral transformation. The computation times for the various steps will be compared for a nontrivial example in Sec. IV. The use of symmetry is straightforward and will not be discussed here.

III. SIMULTANEOUS OPTIMIZATION OF CI COEFFICIENTS AND ORBITALS

Deriving the equations for quadratically convergent orbital optimization we have assumed that the CI coefficients are constant. In general, however, the optimum coefficients depend on the form of the occupied orbitals and vice versa. Thus, after having found a new set of MO's new coefficients and density matrices must be calculated which are used in the next orbital optimization iteration. Since in this alternate process the second derivatives $\partial^2 E / \partial R_{ij} \partial c_I$ are neglected, it is not quadratically convergent. In this section we present a simple extension of the method described in Sec. II which includes the above second derivatives and is therefore quadratically convergent for general MC-SCF wave functions.

Let $\{c^0\}$ be the set of initial CI coefficients and A^0 , B^0 be calculated using these coefficients according to Eqs. (31) and (32), respectively. Then we may write the second order energy expansion in the changes of the orbitals and CI coefficients in the form $(\text{Tr}(A) = \sum_i A_{ii})$

$$E^{(2)} = \sum_{IJ} c_I c_J H_{IJ} + \frac{1}{2} \text{Tr} \{T^* (A^0 + B^0)\} + 2 \sum_I (c_I - c_I^0) \text{Tr} \{T^* A^I\} \quad (51)$$

where $H_{IJ} = \langle \Phi_I | \hat{H} | \Phi_J \rangle$ is the CI matrix and

$$A^I = \frac{1}{2} \left(\frac{\partial A}{\partial c_I} \right)_{(c^0)} \quad (52)$$

The derivative matrices A^I are formally defined as in Eq. (31) but the density matrix derivatives $\gamma^I = \sum_J c_J^0 \gamma^{IJ}$ and $\Gamma^I = \sum_J c_J^0 \Gamma^{IJ}$ have to be used for the calculation of corresponding operators F_{ij}^I . Using the eigenvectors c_I^0 of H_{IJ} we may define $D^I = \sum_J c_J^0 \text{Tr}(T^* A^I)$ and, with the ansatz

$$c_I = \sum_n c_I^n a_n, \quad (53)$$

Eq. (51) becomes

$$E^{(2)} = \sum_n (E^n a_n^2 + 2D^n a_n) - 2D^0 + \frac{1}{2} \text{Tr} \{ \mathbf{T}^* (\mathbf{A}^0 + \mathbf{B}^0) \}, \quad (54)$$

where the E^n are the eigenvalues of H_{IJ} .

Varying this expression with respect to the a_n subject to the normalization condition $\sum_n a_n^2 = 1$ yields the stationary condition

$$\frac{\partial E^{(2)}}{\partial a_n} = 2 \{ E^n a_n + D^n - \epsilon a_n \} = 0 \quad (55)$$

which is easily solved by iterating the equations

$$\epsilon = \frac{\sum_n (E^n a_n^2 + D^n a_n)}{\sum_n a_n^2} \\ a_n = -D^n / (E^n - \epsilon) \quad (56)$$

starting from $a_n^{(1)} = \delta_{0n}$. In case c_I^0 does not correspond to the lowest root of the MC-CI, one should take care that a_0 stays dominant in order to avoid root flipping. Using the new coefficients c_I from (53) the matrices \mathbf{A} and \mathbf{B} are updated as

$$\mathbf{A} = \mathbf{A}^0 + 2 \sum_I (c_I - c_I^0) \mathbf{A}^I \quad (57)$$

$$\mathbf{B} = \mathbf{B}^0 + (\mathbf{A} - \mathbf{A}^0). \quad (58)$$

The second order energy expression (51) then takes the form

$$E^{(2)} = \sum_{IJ} c_I c_J H_{IJ} + \frac{1}{2} \text{Tr} \{ \mathbf{T}^* (\mathbf{A} + \mathbf{B}) \} \quad (59)$$

which is equivalent to Eq. (25). Hence the next iteration step to improve \mathbf{T} can be performed exactly as described in Sec. II. If there is a strong coupling between the orbital rotations and CI coefficients convergence of the above alternate process is sometimes difficult to achieve. In such cases convergence can be much improved if one inverts that part of the second energy derivative matrix which corresponds to rotations between internal orbitals and the CI coefficients (see Appendix B). The additional second energy derivatives needed are given by

$$\left(\frac{\partial^2 E}{\partial c_I \partial R_{ij}} \right)_{\{c^0, R=0\}} = 2(1 - \tau_{ij})(A_{ij}^I - c_I^0 A_{ij}^0) \quad (60)$$

$$\left(\frac{\partial^2 E}{\partial c_I \partial c_J} \right)_{\{c^0, R=0\}} = 2(H_{IJ} - E_0 \delta_{IJ}), \quad (61)$$

where τ_{ij} permutes the indices i and j . Due to the normalization condition in this case the Hessian matrix has at least one zero eigenvalue. Far from convergence the inclusion of the terms (60) may lead to negative eigenvalues of the second derivative matrix. One should therefore start the *simultaneous* optimization of orbitals and coefficients not too far from convergence, e.g., after all parameters R_{ij} are smaller than a certain threshold (0.1–0.01).

The simultaneous optimization of orbitals and CI coefficients only requires the additional calculation of the matrices \mathbf{A}^I . Since no new operators J_{ij} and K_{ij} are

required the additional computational effort is small and in most cases by far overcompensated by the improvement of convergence. This is particularly true if there is a strong coupling between the orbitals and CI coefficients (e.g., if the wave function contains configurations which differ only by single substitutions or if there are linear dependencies between the orbital and CI equations). A striking example for the convergence improvement by the simultaneous optimization of orbitals and CI coefficients is given for a calculation on the CO molecule in Table I, in which 22 configurations within the space of six sigma and two pi orbitals were optimized. Using a delocalized orbital description 12 of these configurations are needed to allow the molecule to dissociate into the two possible ($M_L=0$, $M_L=\pm 1$) asymptotic $1\Sigma^+ (^3P \times ^3P)$ states. The remaining configurations account for electron correlation. The initial orbitals were taken from a preceding Hartree-Fock calculation. In the first column of Table I the orbitals and CI coefficients are optimized alternately. After a rapid initial decrease of the energy, which is mainly due to the optimization of the internal orbitals with respect to the external space, convergence proceeds very slowly and stability to 10^{-6} hartree was not achieved within 20 iterations. The slow convergence is obviously due to a strong coupling of rotations between the valence orbitals and the change of the CI coefficients. In the second column of Table I the simultaneous optimization procedure as described above was applied (started at the fourth iteration). This drastically improves convergence and the required accuracy is now obtained in only five iterations!

Very often the MC-SCF procedure is used to calculate potential curves. Usually the orbitals change only little if one alters the interatomic distance. Therefore one can use the orbitals optimized at a particular point as very suitable start approximation for the optimization at a neighboring point (of course the orbitals must be re-orthonormalized using the new metric \mathbf{S}). An example for such a procedure is given in Table II. In this case 37 configurations for the HF molecule within the space of six sigma and two pi orbitals were optimized. The internuclear distance of the calculation in Table II is 1.9 a.u. and the orbitals optimized at $r_e = 1.73291$ a.u. were employed as start approximation. The extremely fast convergence observed in Table II clearly demonstrates the usefulness of a quadratically convergent optimization procedure for the calculation of potential surfaces.

IV. SIMPLIFIED TREATMENT OF CORE ORBITALS

For the proper description of many phenomena it is sufficient to correlate only a limited number of valence electrons. These move in an average potential of "core orbitals" which are doubly occupied in each determinant of the wavefunction. The core orbitals are usually energetically well separated from the remaining valence and external orbitals and do not change much if one starts the MC-SCF optimization with Hartree-Fock orbitals. Therefore one can introduce approximations into the optimization method described above which reduce the number of operators J_{ij} , K_{ij} , F_{ij} , and G_{ij} required. Clearly, the first energy derivatives $(\partial E / \partial R_{ij})_{R=0}$ must be calculated exactly in order to obtain a correct solution

TABLE I. Convergence of the QC-MCSCF method for the CO molecule^a at $r_e = 2.1318$ a.u.

Iteration	ΔE^b	E'_{\max}	ΔE^c	E'_{\max}	ΔE^d	E'_{\max}
0	-112.803 467 37	0.54 -1	-112.803 467 37	0.54 -1	-112.803 467 37	0.54 -1
1	-0.056 605 91	0.30 -1	-0.056 605 91	0.30 -1	-0.056 667 12	0.30 -1
2	-0.023 787 56	0.31 -1	-0.023 787 56	0.31 -1	-0.023 784 41	0.33 -1
3	-0.007 094 96	0.11 -1	-0.007 094 96	0.11 -1	-0.007 022 55	0.58 -2
4	-0.000 205 80	0.39 -2	-0.000 247 42	0.23 -2	-0.000 261 54	0.27 -2
5	-0.000 002 63	0.38 -2	-0.000 001 92	0.21 -4	-0.000 002 15	0.43 -5
6	-0.000 001 69	0.37 -2	-0.000 000 01	0.76 -5		
7	-0.000 001 61	0.36 -2				
10	-0.000 001 42	0.34 -2				
15	-0.000 001 15	0.31 -2				
20 ^e	-0.000 000 93	0.28 -2				

^a22 configurations see f. Basis set Huzinaga³⁷ 9/5 (6/4); d_{xz} , d_{yz} functions: C(0.7, 0.2), O(1.25, 0.3).

^bWithout simultaneous optimization of orbitals and CI coefficients. 1s orbitals on both atoms treated as core as described in Sec. IV.

^cAs b, but with simultaneous optimization started in the fourth iteration.

^dAs c, but without approximations for core orbitals.

^eNot converged; after 20 iterations the error in the energy is 2.18×10^{-5} a.u. In all cases canonical Hartree-Fock orbitals were used as start approximation; $E_{\text{HF}} = -112.76916795$; final energy: $E_{\text{MC}} = -112.89120514$. In the first iteration Eq. (34) was used, thereafter Eq. (41).

^fIn all configurations the orbitals $1\sigma - 4\sigma$ are doubly occupied. For the remaining electrons we have:

$$^1\Sigma(5\sigma^2, 6\sigma^2, 5\sigma 6\sigma) \times ^1\Sigma(1\pi^4, 1\pi^2 2\pi^2, 2\pi^4) \quad (15 \text{ configurations})$$

$$^1\Sigma(5\sigma 6\sigma) \times ^1\Sigma(1\pi^3 2\pi, 1\pi 2\pi^3) \quad (2 \text{ configurations})$$

$$^3\Sigma(5\sigma 6\sigma) \times ^3\Sigma(1\pi^3 2\pi, 1\pi 2\pi^3, 1\pi^2 2\pi^2) \quad (3 \text{ configurations})$$

$$^1\Sigma(1\pi^4 2\pi^2, 5\sigma^2 6\sigma^2 1\pi^2) \quad (2 \text{ configurations})$$

^gOnly $^3(1\pi_x 2\pi_x)^1(1\pi_y 2\pi_y) + ^1(1\pi_x 2\pi_x)^3(1\pi_y 2\pi_y)$.

as defined by Eq. (35). Approximations to the matrix of second derivatives, however, can only influence the speed of convergence but not the final solution. Although a rigorous quadratic convergence of the optimization process is lost if such approximations are introduced one can hope that convergence remains good if the neglected terms and the corresponding orbital changes are small.

Denoting by $\sum_i^{n_c}$ and $\sum_j^{n_v}$ summations over core and valence orbitals, respectively, the total electronic energy can be written in the form:

$$E_0 = \sum_i^{n_c} \langle i | h + F_c | i \rangle + \sum_{ij}^{n_v} \langle i | F_c \gamma_{ij} + \frac{1}{2} \sum_{kl}^{n_v} \Gamma_{ijkl} J_{kl} | j \rangle, \quad (62)$$

where

$$F_c = h + \sum_k^{n_c} (2J_{kk} - K_{kk}) \quad (63)$$

is the closed shell Fock operator, which describes an effective potential in which the valence electrons are moving. Defining in analogy to Eqs. (26) and (27) the operators F_{ij} and G_{ij} (i, j only corresponding to valence orbitals):

$$F_{ij} = F_c \gamma_{ij} + \sum_{kl}^{n_v} \Gamma_{ijkl} J_{kl} \quad (64)$$

$$G_{ij} = F_{ij} + 2 \sum_{kl}^{n_v} \Gamma_{iklj} K_{kl} \quad (65)$$

and in addition

$$G_c = 2F_c + \sum_{ij}^{n_v} \gamma_{ij} (2J_{ij} - K_{ij}) \quad (66)$$

the second order energy expansion in terms of T takes the form

$$E^{(2)} = E_0 + \frac{1}{2} \text{Tr} \{ T^* (A + B) \} + \sum_{ij}^{n_c} \sum_{kl}^{n_v} T_{kl} \langle k | G_c \delta_{ij} + L_{ij} | 1 \rangle T_{ij} + \sum_{kl}^{n_v} \gamma_{kl} \sum_i^{n_c} \sum_{nm} T_{nk} \langle n | L_{ij} | m \rangle T_{mi}, \quad (67)$$

TABLE II. Computations times^a for the CO calculation in Table I (1s orbitals treated as core).

Step	CPU (sec)	Number of evaluations
Formal matrix elements (γ^{IJ} , Γ^{IJ})	5.7	1
Evaluation of operators J_{ij} , K_{ij}	102.1	7
Transformation of J_{ij} , K_{ij} , h , F_c into orbital basis	42.2	7
Construction and diagonalization of the CI matrix	2.1	7
Evaluation of operators F_{ij} , G_{ij}	5.7	6
Evaluation of matrices A^I	16.7	3
Solution of (non) linear equations	18.5	6
Total	192.8	

^aIBM 370/168 double precision arithmetic. 22 configurations, 124 determinants. Basis set see Table I (48 groups). Only C_{2v} symmetry has been used.

where

$$A_{ij} = \begin{cases} 2\langle i | G_c | j \rangle & \text{if } \phi_j \text{ is core orbital} \\ 2 \sum_k \langle i | F_{jk} | k \rangle & \text{if } \phi_j \text{ is valence orbital} \end{cases} \quad (68)$$

$$B_{ij} = A_{ij} + 2 \sum_k \sum_l \langle i | G_{jk} | l \rangle T_{lk} \quad (69)$$

$$L_{ij} = 2(4K_{ij} - K_{ji} - J_{ij}) \quad (70)$$

Note that $G_{ik} = 0$ if i or k does not correspond to a valence orbital. The first line of Eq. (67) has the same structure as Eq. (25) and completely accounts for all those rotations in which no core orbitals are involved. The last two terms arise from rotations with core orbitals only. In order to simplify Eq. (67) we neglect all off-diagonal second energy derivatives $H_{ijk,kl}$ in which at least one index denotes a core orbital. The corresponding higher order terms are also omitted. This approximation amounts to a first order perturbation treatment and is justified if the core orbitals are energetically well separated from the remaining orbitals. Equation (67) now takes the form

$$E^{(2)} \simeq E_0 + \frac{1}{2} \text{Tr} \{ T_v^* (A_v + B_v) \} + \sum_j \sum_i \{ (A - A^*)_{ij} R_{ij} + \frac{1}{2} D_{ij}^c R_{ij}^2 \} \quad (71)$$

where the matrices T_v , A_v , and B_v have zero elements in those rows and columns which correspond to core orbitals. D_{ij}^c are the diagonal second energy derivatives which account for rotations with core orbitals ϕ_j :

$$D_{ij}^c = 2 \left\{ \langle j | G_{ii} | j \rangle + \langle i | G_c | i \rangle + 2 \langle i | 3K_{jj} - J_{jj} | i \rangle - 2 \sum_k \gamma_{ik} \langle i | 3K_{jj} - J_{jj} | k \rangle \right\} - A_{ii} - A_{jj} \quad (72)$$

In this approximation, only the diagonal operators J_{jj} and K_{jj} are needed for each core orbital. The minimization condition for the space of valence and external orbitals then takes the same form as Eq. (34) (or Eq. (41) if higher order terms are omitted):

$$E' = U_v^* B_v - B_v^* U_v = 0 \quad (73)$$

and can be solved as described in Sec. II. The parameters R_{ij} which describe rotations with core orbitals are simply given by

$$R_{ij} = - (A - A^*)_{ij} / D_{ij}^c \quad (j \leq n_c < i) \quad (74)$$

The convergence of the optimization of the core orbitals can sometimes be improved by transforming the core orbitals as well as the external orbitals such that the operator G_c becomes diagonal in the corresponding subspaces, i.e., takes the "pseudo canonical" form. This transformation clearly does not influence the energy but minimizes the neglected off-diagonal second energy derivatives as far as possible. A further simplification of (72) by completely neglecting the terms involving J_{jj} and K_{jj} does not seem to work satisfactorily: optimization of core orbital space against valence orbital space has been found to proceed too slowly in many cases.

As mentioned above energetically low lying core or-

TABLE III. Convergence of the MC-SCF optimization for the HF molecule; ^a $r = 1.9$ a.u., start with Schmidt-orthonormalized orbitals optimized at $r = 1.7329$ a.u.

Iteration	ΔE	E'_{\max}
0	-100.197 094 86	0.12
1	-0.002 348 76	0.54 - 2
2	-0.000 045 20	0.87 - 4
3	-0.000 000 00	$< 10^{-8}$

^a37 configurations (201 determinants) within the space of 6 σ and 2 π orbitals optimized. A 12s, 7p, 2d, 1f-6s, 2p Gaussian type basis set was used. The Newton-Raphson equations (41) with simultaneous optimization of the orbitals and CI coefficients have been applied.

bitals are not much influenced by the configurational structure of the wave function. Therefore one can also keep them simply fixed as obtained from a preceding Hartree-Fock calculation without much influencing the quality of the wave function. The only additional effort for the core orbitals is then to calculate the operator F_c once. The formulas to be applied for the optimization of the remaining orbitals are the same as in Sec. II with the exception that the one electron Hamiltonian h is replaced by the effective potential F_c . Obviously, F_c does not change during the iteration but has to be transformed into the new orbital basis in each iteration step.

An example for the usefulness of the above approximations is given for the CO molecule in the third column of Table I. In this case the 1s orbitals of both atoms are treated as core orbitals. As expected, quadratic convergence is lost, but the energy converges only slightly slower than in the quadratic case (second column). The computational effort for the various steps of the CO calculation is listed in Table III. Most time is needed to calculate the operators J_{ij} and K_{ij} and for their transformation into orbital basis. Note the small additional effort which is required for the simultaneous optimization of orbitals and CI coefficients (i.e., for the calculation of the matrices A^i).

V. COMPARISON WITH PREVIOUS WORK

In this section we will briefly discuss the relation of our method to previous work, but no attempt is made to review all methods described in the literature. A recent more complete survey is given in Ref. (33).

To show the relation of our quadratically convergent MC-SCF procedure (in the following referred to as QC-MCSCF) to the Fock operator methods we start from Eq. (30). This can be written in the form

$$\sum_j \langle n | F_{mj} | j \rangle + \sum_{jk} \langle n | G_{mj} | k \rangle T_{kj} = \sum_{jk} \epsilon_{jm} \langle n | S | k \rangle U_{kj} \quad (75)$$

or, since this equation must be valid for all vectors $\langle n |$,

$$\sum_j F_{mj} | j \rangle + \sum_{jk} G_{mj} | k \rangle T_{kj} = \sum_{jk} \epsilon_{jm} S | K \rangle U_{kj} \quad (76)$$

If one neglects the exchange operators in the definition

of the operators G_{ij} [Eq. (27)], i. e., approximates $G_{ij} = F_{ij}$ one obtains

$$\sum_{jk} F_{mj} |K\rangle U_{kj} = \sum_{jk} \epsilon_{mj} S |K\rangle U_{kj} \quad (77)$$

or, according to Eq. (20)

$$\sum_j F_{mj} |j'\rangle = \sum_j \epsilon_{mj} S |j'\rangle \quad \text{with } \epsilon_{mj} = \epsilon_{jm}. \quad (78)$$

For $U=1$ these are the well known Hartree-Fock-Roothaan equations which must be satisfied for the optimum MC-SCF wave function. Hence, neglecting the operators K_{kj} in our treatment corresponds to solve the Fock equations for constant F_{ij} , i. e., the functional dependence of the operators F_{ij} from the orbitals is not taken into account.

Using the orthonormality of the new orbitals ϕ'_i , Eq. (78) can also be written in the form

$$\sum_j \langle n' | F_{mj} | j' \rangle - \langle m' | F_{nj} | j' \rangle = 0 \quad (79)$$

which corresponds to Eq. (34) after putting $G_{ij} = F_{ij}$. Hinze¹⁷ has approximated these equations by expanding them up to first order in R and then solved the resulting linear equations. The convergence of this method was not demonstrated, however. Recently, Hinze and Yurtsever¹⁸ have proposed a Jacobi like rotation technique for the solution of Eq. (78). Many papers are concerned with the construction of coupling operators such that Eq. (78) is transformed into pseudo eigenvalue equations from which the orbitals can be obtained in an iterative sequence of matrix diagonalizations. The drawbacks of these methods have already been mentioned in Sec. I. A further source for poor convergence becomes clear from the above derivation of Eq. (78): Obviously, it is exactly valid only for the desired solution. Near convergence it is only a good approximation if the neglected terms in the second energy derivatives are small. For internal-external rotations, the neglected two electron

TABLE V. Comparison of the QC-MCSCF method with the Fock operator method for the BH molecule, ^a $r=3.0$ a.u.

Iteration	ΔE^b	E'_{\max}	ΔE^c	E'_{\max}
0	-25.094 028 23	0.23 -1	-25.094 028 23	0.23 -1
1	-0.012 708 03	0.18 -1	-0.011 400 41	0.19 -1
2	-0.001 049 54	0.94 -2	-0.000 651 64	0.13 -1
3	-0.000 396 76	0.79 -3	-0.000 114 47	0.13 -1
4	-0.000 001 91	0.29 -4	-0.000 092 54	0.12 -1
5			-0.000 086 35	0.12 -1
10			-0.000 066 97	0.11 -1
20			-0.000 041 63	0.92 -2
50			-0.000 011 05	0.22 -2
100			-0.000 001 24	0.16 -2

^aFour configurations, see text. Start with localized Hartree-Fock orbitals; $E_{HF} = -25.084 522 95$, $E_{MC} = -25.108 184 48$.

^bQC-MCSCF, (Eq. 34) was applied in the first iteration, thereafter Eq. (41).

^cWith approximation $G_{ij} = F_{ij}$, i. e., Eq. (79) has been solved. In both cases the orbitals and CI coefficients were optimized simultaneously except for the first iteration. Basis set see legend of Table IV.

integrals are of type $(nk|lm)$, where n, m correspond to external and k, l to internal orbitals. These integrals are of exchange type and can indeed assumed to be small. This seems to be the reason why the closed-shell Hartree-Fock method, which only involves internal-external rotations, is usually well convergent. [In this case all F_{ij} can be represented by one operator F_c and the internal orbitals may be transformed into each other such that the matrix ϵ becomes diagonal. Eq. (78) then becomes a simple pseudo eigenvalue equation. The neglected second order terms are given by the second line of Eq. (67)]. In general, however, the wave function must be optimized also with respect to internal-internal rotations. In this case the neglected integrals in the second energy derivatives involve up to four internal orbitals and may represent large Coulomb type integrals. Their neglect in the minimization condition (79) is probably the reason for poor convergence in many cases.

In order to illustrate these considerations we have tested approximation (78) for a simple MC-SCF calculation on the BH molecule. The results are presented in Tables IV and V. In the case of Table IV we have included all 10 configurations which can be constructed in the space of 4 sigma orbitals. In this case no internal-internal rotations have to be performed since these are accounted for by the freedom in configuration space. The first column of Table IV demonstrates the very fast convergence of QC-MCSCF. In the second column Eq. (34) was solved with the operators G_{ij} replaced by F_{ij} . As expected from the above, the convergence behavior is no longer quadratic but nevertheless not unsatisfactory. In Table V the same comparison is made for a wave function in which only the Hartree-Fock determinant and the three possible double excitations are included. This ansatz yields the same final energy as in Table IV but also internal-internal orbital rotations must be performed. One observes that the convergence of QC-MCSCF is only slightly slower than in Table IV. Neglecting the exchange operators, however, the final solution

TABLE IV. Comparison of the QC-MCSCF method with the Fock operator method for the BH molecule, $r=3.0$ a.u.^a

Iteration	ΔE^b	E'_{\max}	ΔE^c	E'_{\max}
0	-25.096 102 71	0.28 -1	-25.096 102 71	0.28 -1
1	-0.011 769 10	0.20 -1	-0.011 441 82	0.22 -1
2	-0.000 312 30	0.96 -3	-0.000 613 94	0.34 -2
3	-0.000 000 36	0.59 -6	-0.000 023 44	0.46 -3
4			-0.000 002 17	0.15 -3
5			-0.000 000 33	0.63 -4
6			-0.000 000 06	0.27 -4

^a10 configurations (see text). Start with localized Hartree-Fock orbitals; $E_{HF} = -25.084 522 95$, $E_{MC} = -25.108 184 48$.

^bQC-MCSCF [Eq. (34) has been applied].

^cWith approximation $G_{ij} = F_{ij}$, i. e., the Fock equations (79) have been solved. In both cases the orbitals and CI coefficients were optimized simultaneously except for the first iteration. The following Gaussian type basis set was employed: B: s-function exponents (contraction coefficients in brackets): 821.59(0.0586), 124.2(0.4307), 28.38(1.864), 7.893, 2.4213, 0.3381, 0.1077; p_x functions: 2.5865, 0.5197, 0.1235; d_{x^2} functions: 1.0, 0.5; H: s functions: 4.5004, 0.6813, 0.1514; p_x function: 0.5.

was not obtained within 100 iterations, although the energy decreased monotonically!

Recently, Golebiewski, Hinze, and Yurtsever³⁵ have proposed the so called "orthogonal gradient" method, in which the operators G_{ij} are completely neglected, i.e., an even more drastic approximation of the second energy derivatives is made. This leads to a similar minimization condition as Eq. (34), but the matrix B , which depends on T , is replaced by the constant matrix A . The authors found their method well convergent in some semi-empirical test calculations. We have tried it in some *ab initio* calculations but achieved only very slow convergence with a strong damping [as described in Ref. (35)] of the process. As to the efficiency of these Fock operator methods based on the F_{ij} only, we may say the following: since the neglect of the operators K_{ij} reduces the effort per iteration by about a factor of three only (the J_{ij} are required in any case), these methods should converge within 8–12 iterations in order to be competitive. This seems quite unlikely in the general case.

Various orbital optimization techniques have been proposed which are based on the direct minimization of the energy.^{25–34} The exponential ansatz $U = e^{\mathbf{R}}$ has been used in the MC-SCF context for the first time by Levy.²⁵ He has approximated the second energy derivatives by a finite difference method. This requires to calculate the operators F_{ij} a large number of times and seems not to be very efficient. Several authors^{30–34} have expanded the energy up to second order in \mathbf{R} and solved the resulting Newton-Raphson equations. These methods, which actually have much inspired the present work, have already been discussed in Sec. II. Yaffe and Goddard³⁰ have also derived equations for cubic convergence, i.e., included all third order terms of \mathbf{R} in the energy expansion. This requires, however, nearly a full transformation of the two electron integrals in each iteration step which is very computation time consuming. We note that the internal set of operators J_{ij} and K_{ij} is sufficient to calculate the energy derivatives of any order as far as only internal-internal rotations are considered. Thus, it would in principle be possible to optimize the internal orbital space exactly in each iteration. Since the resulting equations are much more complicated than those to be solved in the QC-MCSCF method such a procedure has not been developed so far.

A third group of MC-SCF methods^{20–23} is based on the generalized Brillouin theorem derived by Levy and Berthier.²⁴ This theorem states that for the optimum orbitals the Hamilton matrix elements between the MC-SCF wave function and all singly excited configurations vanish:

$$\langle \Psi_{ij} | \hat{H} | \Psi \rangle = 0, \quad (80)$$

where

$$\Psi_{ij} = (1 - \tau_{ij})(\eta_i^{\alpha*} \eta_j^{\alpha} + \eta_i^{\beta*} \eta_j^{\beta}) \Psi. \quad (81)$$

$\eta_i^{\alpha*}$ and η_j^{β} are the usual creation and annihilation operators, respectively, which act on spin orbitals $\phi_i, \alpha, \phi_j, \beta$. τ_{ij} permutes the indices i and j . One may now define an operator \hat{R}

$$\hat{R} = \sum_{ij} R_{ij}(\eta_i^{\alpha*} \eta_j^{\alpha} + \eta_i^{\beta*} \eta_j^{\beta}), \quad R_{ij} = -R_{ji} \quad (82)$$

such that $\hat{R}\Psi$ is a linear combination of all Brillouin-singly excited configurations (SC's) with coefficients R_{ij} . To first order, the action of $(1 + \hat{R})$ on Ψ has the same effect as performing the orbital transformation

$$\mathbf{C}' = \mathbf{C}(1 + \mathbf{R}). \quad (83)$$

Therefore it is possible to determine the orbital rotation parameters R_{ij} in a CI calculation in the space of the MCSCF reference function plus all SC's (SC-CI) and then "absorb" the SC's according to Eq. (83) approximately into the reference function. Since the above orbital transformation preserves orthonormality only to first order the new orbital vectors \mathbf{C}' must be reorthonormalized.²⁰ Alternatively, the new set of orbitals can be derived from the natural orbitals of the SC-CI.²³ If the above process is repeated and converges, the matrix elements (80) finally vanish and, hence, the coefficients R_{ij} (this may easily be seen from first order perturbation theory). Actually, it is easy to show that

$$\left(\frac{\partial E}{\partial R_{ij}} \right)_{\mathbf{R}=0} = 2 \langle \Psi_{ij} | \hat{H} | \Psi \rangle \quad (84)$$

and that Eqs. (80) and (35) are equivalent. The generalized Brillouin theorem MC-SCF method (GBT-MCSCF) seems to be rather stable with respect to convergence and has been successfully applied to various calculations of ground and excited molecular states. A simultaneous optimization of the orbitals and CI coefficients has been proposed by Chang and Schwarz.²² It consists simply of varying in the SC "super CI" the MC-SCF coefficients simultaneously with the coefficients R_{ij} of the SC's (note that the SC's themselves depend on the CI coefficients, which are taken from the preceding iteration). However, the GBT-MCSCF method has the following disadvantages: First, rather large Hamilton and overlap matrices must be constructed in each iteration. This step is not trivial. One can show, that the information needed to construct the Hamilton matrix is provided by the spin-free reduced density matrices up to fourth order for the MC-SCF reference function (see Appendix C). Second, the GBT-MCSCF method does not converge quadratically, although it is more complicated than QC-MCSCF and requires the same transformation of the two electron integrals in each iteration. The relation of the GBT-MCSCF method to the quadratically convergent Newton-Raphson method can be shown by defining in analogy to Eq. (22) the unitary operator

$$\hat{U} = e^{\hat{R}} = 1 + \hat{R} + \frac{1}{2} \hat{R}^2 + \dots \quad (85)$$

and expanding the energy up to second order in \hat{R} :

$$\begin{aligned} E &= \langle \hat{U}\Psi | \hat{H} | \hat{U}\Psi \rangle \\ &= \langle \Psi | \hat{H} | \Psi \rangle + 2 \langle \hat{R}\Psi | \hat{H} | \Psi \rangle + \langle \hat{R}^2\Psi | \hat{H} | \Psi \rangle \\ &\quad + \langle \hat{R}\Psi | \hat{H} | \hat{R}\Psi \rangle + O(\mathbf{R}^3). \end{aligned} \quad (86)$$

Minimization of this second order energy approximation corresponds exactly to the Newton-Raphson method (see Appendix C). However, solving the SC super CI eigenvalue equation corresponds to the minimization of the following variational energy expression:

TABLE VI. Comparison of the QC-MCSCF and GBT-MCSCF methods for the BH molecule, $r=3.0$ a.u.^a

Iteration	ΔE (QC-MCSCF)	μ^b	ΔE (GBT-MCSCF)	μ
0	-25.092 414 92	0.1779	-25.092 414 92	0.1779
1	-0.013 497 50	0.2397	-0.013 201 51	0.2424
2	-0.001 354 84	0.2482	-0.001 478 57	0.2407
3	-0.000 037 69	0.2332	-0.000 152 94	0.2367
4	-0.000 000 00	0.2332	-0.000 038 86	0.2367
5			-0.000 011 94	0.2356
6			-0.000 003 99	0.2346
7			-0.000 001 41	0.2341
8			-0.000 000 52	0.2337
9			-0.000 000 19	0.2335
10			-0.000 000 07	0.2334

^aTwo configurations: $1\sigma^2 2\sigma^2 3\sigma^2$, $1\sigma^2 2\sigma^2 4\sigma^2$. Basis set see Table IV. Start with localized Hartree-Fock orbitals.

^bDipole moments in atomic units (1 a.u. = 2.5417 D).

$$E = \langle \Psi | \hat{H} | \Psi \rangle + 2 \langle \hat{R} \Psi | \hat{H} | \Psi \rangle + \langle \hat{R} \Psi | \hat{H} - E | \hat{R} \Psi \rangle \quad (87)$$

which differs from the second order approximation by the terms

$$- \langle \hat{R}^2 \Psi | \hat{H} | \Psi \rangle + \langle \hat{R} \Psi | \hat{R} \Psi \rangle E. \quad (88)$$

The neglect of these terms destroys the quadratic convergence and, at the same time, makes the energy expression much more complicated since the terms $\langle \hat{R}^2 \Psi | \hat{H} | \Psi \rangle$ and $\langle \hat{R} \Psi | \hat{H} | \hat{R} \Psi \rangle$ largely cancel each other in (86) but not so in (87) (see Appendix C). On the other hand, the stability of the GBT-MCSCF method with respect to convergence probably results from the last term in Eq. (88), which accounts for the normalization condition. Chang and Schwarz²² have tried to include the neglected terms in the super CI by modifying the elements of the SC-CI Hamilton matrix. Their idea was to minimize the energy expectation value of a configuration expansion which corresponds to second order in \hat{R} to the orbital transformation.²² (According to Eq. (83) the usual SC-CI expansion is only correct to first order in \hat{R}). In fact, the unitary orbital transformation (22) is equivalent to the configuration expansion

$$\Psi' = (1 + \hat{R} + \frac{1}{2} \hat{R}^2 + \dots) \Psi, \quad (89)$$

which, using the usual anticommutation rules for the creation and annihilation operators (see Appendix C), may be written in the more familiar form

$$\begin{aligned} \Psi' = \Psi + \sum_{ij} R_{ji} \Psi_i^j + \frac{1}{2} \sum_{ijkl} R_{ji} R_{lk} \Psi_{ik}^{jl} \\ + \frac{1}{2} \sum_{ijk} R_{jk} R_{ki} \Psi_i^j + O(R^3), \end{aligned} \quad (90)$$

where

$$\Psi_i^j = (\eta_j^{\alpha*} \eta_i^{\alpha} + \eta_j^{\beta*} \eta_i^{\beta}) \Psi \quad (91)$$

$$\begin{aligned} \Psi_{ik}^{jl} = (\eta_j^{\alpha*} \eta_i^{\alpha*} \eta_k^{\alpha} \eta_i^{\alpha} + \eta_j^{\alpha*} \eta_i^{\beta*} \eta_k^{\beta} \eta_i^{\alpha} \\ + \eta_j^{\beta*} \eta_i^{\alpha*} \eta_k^{\alpha} \eta_i^{\beta} + \eta_j^{\beta*} \eta_i^{\beta*} \eta_k^{\beta} \eta_i^{\beta}) \Psi. \end{aligned} \quad (92)$$

Up to second order in \hat{R} , the resulting energy expectation value is identical with equation (86). By comparing (90) with the equation (6) of Chang and Schwarz²² it appears that in the latter the second order terms are in-

cluded incompletely. For comparison of convergence behavior, we have performed a GBT-MCSCF calculation for the BH molecule with a two configuration wave function. Since in this simple case the first order reduced density matrix (in MO basis) is diagonal, the improved orbitals after each iteration step can be obtained as the most strongly occupied natural orbitals of the SC-CI wavefunction (i.e., by diagonalizing the corresponding reduced first order density matrix as proposed in Ref. (23)). The convergence of this procedure is compared to QC-MCSCF in Table VI. In both methods the orbitals and CI coefficients have been optimized simultaneously. Far from convergence both methods proceed about equally well. However, after all orbital rotations are small, QC-MCSCF terminates rapidly, whereas the GBT-MCSCF method only slowly approaches the final result. Although one could argue that the energy changes after the fourth or fifth iteration are negligible, one should keep in mind that errors in the energy are of second order. The convergence of other properties, as for instance dipole moments, which are also given in Table VI, can be much worse. The advantage of a quadratically convergent procedure in this respect is obvious. The advantage of QC-MCSCF for calculating energy surfaces has already been demonstrated in Sec. III.

We may conclude, that the quadratically convergent orbital optimization scheme as presented in this paper in connection with the simultaneous optimization of the CI coefficients seems to be one of the most efficient (with respect to convergence and computation time) and simplest (with respect to the formalism and programming) MC-SCF methods presently available.

ACKNOWLEDGMENTS

We like to thank Professor E. A. Reinsch for many stimulating discussions and continuous support during the development for the method. The excellent computer facilities provided by the Hochschulrechenzentrum der Universität Frankfurt and the Hochschulrechenzentrum der Technischen Hochschule Darmstadt are acknowledged.

APPENDIX A. ALGORITHM FOR THE EVALUATION OF THE OPERATORS J_{ij} AND K_{ij}

In the following we assume the two electron integrals $(\mu\nu|\rho\sigma)$ to be given in canonical order ($\mu \geq \nu$), ($\rho \geq \sigma$), $\{\mu\nu\} \geq \{\rho\sigma\}$, where $\{\mu\nu\} = \mu(\mu-1)/2 + \nu$. The integrals are stored on an external device and transferred blockwise into a high speed memory buffer. The orbital vectors are stored one after another in the array C , which may be written for convenience as $C(\mu, i)$. The first index specifies the basis functions, the second the orbitals. The elements of the operators $J_{ij}(\mu, \nu)$ and $K_{ij}(\mu, \nu)$ are also stored lineary. Since the operators J_{ij} are symmetric, only the lower triangles need to be stored such that the position of the element $J_{ij}(\mu, \nu)$ is $\{\mu\nu\}$. A simplified form of the program is given in Fig. 1. In the scheme N is the number of basis functions (AO's) and M the number of different orbitals needed to calculate the operators. The optimum form of

the algorithm requires to evaluate all operators simultaneously during one sequential read of the integral tape which is possible in most cases. If not all Coulomb and exchange operators can be kept simultaneously in high speed memory, the Coulomb and exchange operators can be calculated separately without much loss of efficiency (instead of $3/8N^4M$ operations in the innermost loop then $1/2N^4M$ operations are required). If even this is not possible, only a smaller number of operators is calculated per integral read and, hence, the calculations in the innermost loop have to be repeated. It should be noted, that this loop is often not rate determining, such that the loss of efficiency is not severe. The use of symmetry is straightforward and not considered here (if molecular symmetry can be used the operators are blocked and only the same blocks of operators of the same symmetry type must be evaluated simultaneously). The number of multiplications for the evaluation of $M(M+1)/2$ operators J_{ij} and K_{ij} is about (terms proportional to N^2 and N being neglected)

$$Z_1 = (3/8N^4 + \frac{3}{4}N^3)M + \frac{7}{8}N^3M(M+1). \quad (A1)$$

The transformation of the operators into MO basis requires about $7/4N^3M(M+1)$ additional multiplications. The total number

$$Z = 3/8N^4M + 35/12N^3M^2 + 11/3N^3M \quad (A2)$$

can be compared to two alternate algorithms proposed by Cheung, Elbert and Ruedenberg.²³ These require about

$$Z_a = N^4M + \frac{3}{2}N^3M^2 + \frac{5}{2}N^3M \quad (A3)$$

$$Z_b = \frac{1}{2}N^4M + \frac{9}{4}N^3M^2 + \frac{3}{2}N^3M \quad (A4)$$

operations. Algorithm (A3) is applied in the GBT-MCSCF program developed by the above authors. As an example, we compare the three algorithms for $N=30$ and $M=6$. Our method requires about 5.2×10^6 multiplications, algorithm (A3) 6.7×10^6 and (A4) 4.9×10^6 operations. If the Coulomb and exchange operators are evaluated separately, our method requires about 5.9×10^6 multiplications.

Program for the evaluation of the operators J_{ij} and K_{ij}

The integrals $(\mu\nu|\rho\sigma)$ are assumed to be premultiplied by the factors $(2 - \delta_{\mu\nu})(2 - \delta_{\rho\sigma})(2 - \delta_{\mu\nu\rho\sigma})$.

Initialize all operators J_{ij} and K_{ij} to zero.

$\mu\nu=1$

DO 100 $\mu=1, N$

$T^\nu(\rho\sigma, i)=0$ for $1 \leq \rho\sigma \leq \{\mu, \mu\}$, $1 \leq i \leq M$

DO 200 $\nu=1, \mu$

$T^\rho(\rho, i)=0$

$T^\sigma(\rho, i)=0$ for $1 \leq \rho \leq \mu$, $\mu \leq i \leq M$

$\rho\sigma=1$

DO 300 $\rho=1, \mu$

$\sigma_\rho=\rho$

IF $(\mu \cdot EQ \cdot \rho) \sigma_\rho = \nu$

DO 400 $\sigma=1, \sigma_\rho$

DO 500 $i=1, M$

$T^\sigma(\rho, i) = T^\sigma(\rho, i) + (\mu\nu|\rho\sigma) C(\sigma, i)$

$T^\rho(\sigma, i) = T^\rho(\sigma, i) + (\mu\nu|\rho\sigma) C(\rho, i)$

500 $T^\nu(\rho\sigma, i) = T^\nu(\rho\sigma, i) + (\mu\nu|\rho\sigma) C(\nu, i)$

400 $\rho\sigma = \rho\sigma + 1$

300 CONTINUE

$T^\rho(\sigma, i) = T^\rho(\sigma, i) + T^\sigma(\sigma, i)$ for $1 \leq \sigma \leq \mu$, $1 \leq i \leq M$

for all operators K_{ij} :

DO 250 $\sigma=1, \mu$

$K_{ij}(\sigma, \nu) = K_{ij}(\sigma, \nu) + T^\rho(\sigma, i) C(\mu, j)$

$K_{ij}(\nu, \sigma) = K_{ij}(\nu, \sigma) + T^\rho(\sigma, j) C(\mu, i)$

$K_{ij}(\sigma, \mu) = K_{ij}(\sigma, \mu) + T^\rho(\sigma, i) C(\nu, j)$

250 $K_{ij}(\mu, \sigma) = K_{ij}(\mu, \sigma) + T^\rho(\sigma, j) C(\nu, i)$

for all operators J_{ij} :

DO 260 $\rho=1, \mu$

260 $J_{ij}(\mu\nu) = J_{ij}(\mu\nu) + T^\sigma(\rho, i) C(\rho, j) + T^\sigma(\rho, j) C(\rho, i)$

200 $\mu\nu = \mu\nu + 1$

for all operators J_{ij} :

DO 150 $\rho\sigma=1, \{\mu, \mu\}$

150 $J_{ij}(\rho\sigma) = J_{ij}(\rho\sigma) + T^\nu(\rho\sigma, i) C(\mu, j) + T^\nu(\rho\sigma, j) C(\mu, i)$

100 CONTINUE

for all operators J_{ij} :

Multiply diagonal elements by two.

APPENDIX B. CONVERGENCE IMPROVEMENT OF THE ITERATIVE SOLUTION OF THE (NON)-LINEAR EQUATIONS

If there is a strong coupling between the orbital rotations or, what happens more often, between the orbital changes and the CI-coefficients, convergence of the iterative algorithm for the solution of the (non)-linear equations as described in Secs. II and III can be much improved by inverting that part of the Hessian matrix which corresponds to rotations between internal orbitals and the CI coefficients. Let the linear equations to be partitioned as

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} + \begin{pmatrix} \mathbf{H}_{aa} & \mathbf{H}_{ab} \\ \mathbf{H}_{ab}^* & \mathbf{H}_{bb} \end{pmatrix} \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix} = 0 \quad (B1)$$

and let $\mathbf{x}_a^{(i)}$, $\mathbf{x}_b^{(i)}$ the i th approximation to their solution:

$$\begin{pmatrix} \mathbf{y}_a^{(i)} \\ \mathbf{y}_b^{(i)} \end{pmatrix} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} + \begin{pmatrix} \mathbf{H}_{aa} & \mathbf{H}_{ab} \\ \mathbf{H}_{ab}^* & \mathbf{H}_{bb} \end{pmatrix} \begin{pmatrix} \mathbf{x}_a^{(i)} \\ \mathbf{x}_b^{(i)} \end{pmatrix}. \quad (B2)$$

Assuming that the elements of the matrix \mathbf{H}_{ab} are small we may improve $\mathbf{x}_a^{(i)}$ and $\mathbf{x}_b^{(i)}$ independently:

$$\mathbf{y}_a^{(i)} + \mathbf{H}_{aa} \Delta \mathbf{x}_a^{(i)} = 0 \quad (B3a)$$

$$\mathbf{y}_b^{(i)} + \mathbf{H}_{bb} \Delta \mathbf{x}_b^{(i)} = 0. \quad (B3b)$$

Assuming that the off-diagonal elements of the matrix H_{bb} are small as compared to the diagonal elements, which is mostly the case if \mathbf{x}_b corresponds to internal-external rotations only, Eq. (B3b) can be solved approximately as described in Sec. II Eq. (B3a) can be solved by diagonalizing H_{aa} :

$$(UH_{aa}U^*)_{ij} = D_{ii} \delta_{ij} \quad (B4)$$

$$(\Delta x_a)_i = \sum_{kl} U_{li} U_{lk} (y_a)_k / D_{li} \quad (B5)$$

The rows and columns of H_{aa} which correspond to redundant parameters should be eliminated or set equal to zero prior to the diagonalization. In Eq. (B5) those summands for which $D_{li} = 0$ have to be omitted. Non-linear terms can be included in the calculation of $y_a^{(i)}$ and $y_b^{(i)}$.

In order to avoid divergence if the Hessian matrix is not positive definite in ground state calculations the denominators D_{li} should all be set positive and sufficiently large. [If linear equations are solved the terms for which D_{li} has been altered are included only in the first iteration in Eq. (B5).] However, if the n th excited state is optimized and the orbitals and CI coefficients are optimized simultaneously, the Hessian matrix has $n-1$ negative eigenvalues. Clearly, these may not simply set positive as one should do for those additional negative eigenvalues which may arise from orbital rotations. In order to distinguish between the two types a further partitioning of the matrix H_{aa} is performed:

$$H_{aa} = \begin{pmatrix} H_{ij,kl} & H_{ij,j} \\ H_{i,j} & D_{i,j} \end{pmatrix}, \quad (B6)$$

where $H_{ij,kl} = (\partial^2 E / \partial R_{ij} \partial R_{kl})$, $H_{ij,j} = (\partial^2 E / \partial R_{ij} \partial c_j)$ and $D_{i,j} = (\partial^2 E / \partial c_i \partial c_j)$ [see Eqs. (43), (60)–(61)]. The orbital rotation parameters are obtained by solving

$$y_{ij}^{(i)} + \sum_{k>i} \tilde{H}_{ij,kl} \Delta R_{kl}^{(i)} = 0, \quad (B7)$$

where

$$\tilde{H}_{ij,kl} = H_{ij,kl} - \sum_{IJ} H_{ij,I} D_{I,J}^{-1} H_{J,kl}. \quad (B8)$$

The matrix D^{-1} can be evaluated from the eigenvectors \mathbf{c}^k and eigenvalues E_k of the CI matrix H_{IJ} , which are known from the preceding MC-CI calculation:

$$D_{IJ}^{-1} = \sum_{k \neq 0} c_I^k (E_k - E_0)^{-1} c_J^k. \quad (B9)$$

Eq. (B7) is solved together with Eq. (B3b) as described above with the restriction that all denominators are positive.

APPENDIX C. MATRIX ELEMENTS BETWEEN BRILLOUIN-SINGLY-EXCITED CONFIGURATIONS

In second quantization the Hamilton operator takes the form

$$\hat{H} = \sum_{ij} \langle i|h|j \rangle (\eta_i^{\alpha*} \eta_j^{\alpha} + \eta_i^{\beta*} \eta_j^{\beta})$$

$$+ \frac{1}{2} \sum_{ijkl} (ij|kl) (\eta_i^{\alpha*} \eta_k^{\alpha*} \eta_j^{\alpha} \eta_l^{\alpha} + \eta_i^{\alpha*} \eta_k^{\beta*} \eta_j^{\beta} \eta_l^{\alpha} + \eta_i^{\beta*} \eta_k^{\alpha*} \eta_j^{\alpha} \eta_l^{\beta} + \eta_i^{\beta*} \eta_k^{\beta*} \eta_j^{\beta} \eta_l^{\beta}) \\ = \hat{h} + \hat{g} \quad (C1)$$

The energy expectation value for the MC-SCF reference function Ψ is given by

$$E_0 = \langle \Psi | \hat{H} | \Psi \rangle \\ = \sum_{ij} \langle i|h|j \rangle \gamma_{ji} + \frac{1}{2} \sum_{ijkl} (ij|kl) \Gamma_{ij,kl}^{(2)}, \quad (C2)$$

where

$$\langle \Psi | \Psi \rangle = 1 \quad (C3)$$

$$\gamma_{ij} = \langle \Psi | \eta_i^{\alpha*} \eta_j^{\alpha} + \eta_i^{\beta*} \eta_j^{\beta} | \Psi \rangle \quad (C4)$$

$$\Gamma_{ij,kl}^{(2)} = \langle \Psi | \eta_i^{\alpha*} \eta_k^{\alpha*} \eta_j^{\alpha} \eta_l^{\alpha} + \eta_i^{\alpha*} \eta_k^{\beta*} \eta_j^{\beta} \eta_l^{\alpha} + \dots | \Psi \rangle. \quad (C5)$$

According to this definition $\Gamma^{(2)}$ obeys the relations

$$\Gamma_{ij,kl}^{(2)} = -\Gamma_{kl,ij}^{(2)} = -\Gamma_{li,jk}^{(2)} = \Gamma_{ji,ik}^{(2)}. \quad (C6)$$

Note that the quantity Γ as introduced in Sec. II is symmetrized in the index pairs ij and kl and is related to $\Gamma^{(2)}$ as follows:

$$\Gamma_{ij,kl} = \frac{1}{2} (\Gamma_{ij,kl}^{(2)} + \Gamma_{ji,kl}^{(2)}) = \frac{1}{2} (\Gamma_{ij,kl}^{(2)} + \Gamma_{ij,lk}^{(2)}). \quad (C7)$$

The energy expectation value for the SC-CI wave function can be written in the form

$$E = E_0 + 2 \langle \Psi | \hat{H} \hat{R} | \Psi \rangle + \langle \Psi | \hat{R}^* (\hat{H} - E) \hat{R} | \Psi \rangle, \quad (C8)$$

where the operator \hat{R} is given by Eq. (82). Hence, in order to perform the SC-CI calculation, the matrix elements $\langle \Psi | \hat{H} \hat{R} | \Psi \rangle$, $\langle \Psi | \hat{R}^* \hat{R} | \Psi \rangle$ and $\langle \Psi | \hat{R}^* \hat{H} \hat{R} | \Psi \rangle$ must be evaluated. Using the anticommutation rules

$$[\eta_i^{\mu}, \eta_j^{\nu}]_{\pm} = [\eta_i^{\mu*}, \eta_j^{\nu*}]_{\pm} = 0 \quad \mu, \nu = \{\alpha, \beta\} \quad (C9)$$

$$[\eta_i^{\mu*}, \eta_j^{\nu}]_{\pm} = \delta_{ij} \delta_{\mu\nu} \quad (C10)$$

and the relation $R_{ij} = -R_{ji}$ and Eqs. (12), (C6) one derives:

$$\langle \Psi | \hat{R} | \Psi \rangle = \sum_{ij} R_{ij} \gamma_{ji} = 0 \quad (C11)$$

$$\langle \Psi | \hat{R}^* \hat{R} | \Psi \rangle = \sum_{ij} \sum_{kl} R_{ji} R_{kl} (\Gamma_{ij,kl}^{(2)} + \delta_{jk} \gamma_{li}) \quad (C12)$$

$$\langle \Psi | \hat{h} \hat{R} | \Psi \rangle = \sum_{ij} \gamma_{ij} \langle i|h|j \rangle R_{nj} \quad (C13)$$

$$\langle \Psi | \hat{g} \hat{R} | \Psi \rangle = \sum_{ij} \sum_{kl} \sum_n \Gamma_{ij,kl}^{(2)} \langle i|J_{kl}|n \rangle R_{nj} \quad (C14)$$

$$\langle \Psi | \hat{R}^* \hat{h} \hat{R} | \Psi \rangle = \sum_{ij} \sum_{kl} \sum_{mn} \langle i|h|j \rangle R_{lk} R_{mn} (\Gamma_{ij,kl}^{(3)} + 2\delta_{jm} \Gamma_{in,kl}^{(2)} + \delta_{im} \Gamma_{kn,ij}^{(2)} + \delta_{il} \delta_{jm} \gamma_{kn}) \quad (C15)$$

$$\langle \Psi | \hat{R}^* \hat{g} \hat{R} | \Psi \rangle = \frac{1}{2} \sum_{ij} \sum_{kl} \sum_{mn} \sum_{ab} (ij|kl) R_{nm} R_{ab} (\Gamma_{ij,kl}^{(4)} + \delta_{ab} \gamma_{mn} + 4\delta_{ja} \Gamma_{ib,mn,kl}^{(3)} + \delta_{na} \Gamma_{mb,ij,kl}^{(3)} + 2\delta_{ja} \delta_{in} \Gamma_{mb,kl}^{(2)} + 2\delta_{ia} \delta_{nl} \Gamma_{mj,kb}^{(2)}). \quad (C16)$$

The third and fourth order reduced density matrices $\Gamma^{(3)}$ and $\Gamma^{(4)}$ respectively, are defined as

$$\Gamma_{ij,kl,mn}^{(3)} = \langle \Psi | \eta_i^{\alpha*} \eta_k^{\alpha*} \eta_m^{\alpha*} \eta_n^{\alpha} \eta_i^{\alpha} \eta_j^{\alpha} + \eta_i^{\beta*} \eta_k^{\beta*} \eta_m^{\beta*} \eta_n^{\beta} \eta_i^{\beta} \eta_j^{\beta} + \dots | \Psi \rangle \quad (C17)$$

$$\Gamma_{ij,kl,mn,ab}^{(4)} = \langle \Psi | \eta_i^{\alpha*} \eta_k^{\alpha*} \eta_m^{\alpha*} \eta_a^{\alpha*} \eta_b^{\alpha} \eta_i^{\alpha} \eta_j^{\alpha} + \dots | \Psi \rangle. \quad (C18)$$

(The omitted summations over different spin components are obvious; the spins in pairs (ij) , (kl) , (mn) , and (ab) are always the same, i.e., there are 8 spin components for $\Gamma^{(3)}$ and 16 for $\Gamma^{(4)}$.)

From the above formulae, which involve the density matrices up to fourth order, it is clear that the SC-CI matrix is much more difficult to calculate than the second energy derivative matrix given by Eq. (43). Since in all density matrices in the above formulae at least two of the indices i, j, k or l occur the same set of operators J_{ij} and K_{ij} is needed to calculate both matrices. Actually, the complicated structure of the SC-CI matrix is due to the neglect of the terms $\langle \Psi | \hat{H} \hat{R} \hat{R} | \Psi \rangle$ in the energy expansion. For these matrix elements one derives:

$$\langle \Psi | \hat{H} \hat{R} \hat{R} | \Psi \rangle = - \sum_{ij} \sum_{kl} \sum_{mn} \langle i | h | j \rangle R_{ik} R_{mn} (\Gamma_{ij,kl,mn}^{(3)} + 2\delta_{jm} \Gamma_{in,kl}^{(2)} + \delta_{im} \Gamma_{kn,ij}^{(2)} + \delta_{jh} \delta_{mi} \gamma_{in}) \quad (C19)$$

$$\langle \Psi | \hat{g} \hat{R} \hat{R} | \Psi \rangle = - \frac{1}{2} \sum_{ij} \sum_{kl} \sum_{mn} \sum_{ab} \langle ij | kl \rangle R_{nm} R_{ab} (\Gamma_{ij,kl,mn,ab}^{(4)} + 4\delta_{ja} \Gamma_{ib,mn,kl}^{(3)} + \delta_{na} \Gamma_{mb,ij,kl}^{(3)} + 2\delta_{jm} \delta_{ia} \Gamma_{in,kb}^{(2)} + 2\delta_{jm} \delta_{na} \Gamma_{ib,kl}^{(2)}) \quad (C20)$$

Summing up Eqs. (C13)–(C16) and (C19), (C20) all terms involving the third and fourth order density matrices cancel and one obtains

$$E^{(2)} = E_0 + 2 \sum_{ijn} \langle i | h | j \rangle \gamma_{in} + \sum_{kl} \Gamma_{ij,kl}^{(2)} J_{kl} | n \rangle (R + \frac{1}{2} R^2)_{nj} + \sum_{ij} \sum_{mn} R_{mi} \langle m | h | j \rangle \gamma_{ij} + \sum_{kl} \{ \Gamma_{ij,kl}^{(2)} J_{kl} + (\Gamma_{ik,jl}^{(2)} + \Gamma_{ik,lj}^{(2)}) K_{kl} \} | n \rangle R_{nj}. \quad (C21)$$

This is equivalent to Eq. (25) truncated to second order in R .

¹C. C. J. Roothaan, Rev. Mod. Phys. 32, 179 (1960).

²S. Huzinaga, Phys. Rev. 120, 866 (1960); 122, 131 (1961).

³F. W. Birss and S. Fraga, J. Chem. Phys. 38, 2552 (1963); 40, 3212 (1964); S. Fraga and F. W. Birss, J. Chem. Phys. 40, 3203 (1964); 40, 3207 (1964).

⁴G. Das and A. C. Wahl, J. Chem. Phys. 44, 87 (1966); 47, 2934 (1967); G. Das, J. Chem. Phys. 46, 1568 (1967).

⁵A. Veillard and E. Clementi, Theoret. Chim. Acta (Berl.) 7, 133 (1967).

⁶J. Hinze and C. C. J. Roothaan, Suppl. Prog. Theor. Phys.

40, 37 (1967).

⁷S. Huzinaga, Prog. Theor. Phys. 41, 307 (1969); J. Chem. Phys. 51, 3971 (1969).

⁸W. A. Goddard, T. H. Dunning Jr., and W. J. Hunt, Chem. Phys. Lett. 4, 231 (1969).

⁹G. Das and A. C. Wahl, J. Chem. Phys. 56, 1769 (1972).

¹⁰G. Das, J. Chem. Phys. 58, 5104 (1973).

¹¹K. Hirao and H. Nakatsuji, J. Chem. Phys. 59, 1457 (1973); K. Hirao, J. Chem. Phys. 60, 3215 (1973).

¹²M. H. Wood and A. Veillard, Mol. Phys. 26, 595 (1973).

¹³R. Caballol, R. Gallifa, J. M. Riera, and R. Carbo, Int. J. Quant. Chem. 8, 373 (1974).

¹⁴R. Carbo, R. Gallifa, and J. M. Riera, Chem. Phys. Lett. 30, 43 (1975).

¹⁵J. Hendekevic, Theoret. Chim. Acta (Berl.) 42, 193 (1976).

¹⁶Y. Ishikawa, Chem. Phys. Lett. 37, 597 (1976).

¹⁷J. Hinze, J. Chem. Phys. 59, 6424 (1973).

¹⁸J. Hinze and E. Yurtsever, J. Chem. Phys. 70, 3188 (1979).

¹⁹A. C. Wahl and G. Das, in *Modern Theoretical Chemistry*, edited by H. F. Schaefer, Vol. III, p. 51 (Plenum, New York, 1977).

²⁰F. Grein and T. C. Chang, Chem. Phys. Lett. 12, 44 (1971); F. Grein and A. Banerjee, Int. J. Quant. Chem. Symp. 9, 147 (1975); A. Banerjee and F. Grein, Int. J. Quant. Chem. 10, 123 (1976).

²¹P. A. Ruttnik and J. H. van Lenthe, Theoret. Chim. Acta (Berl.) 44, 97 (1977).

²²T. C. Chang and W. H. E. Schwarz, Theoret. Chim. Acta (Berl.) 44, 45 (1977).

²³K. Ruedenberg, L. M. Cheung, and S. T. Elbert, Int. J. Quant. Chem. 16, 1068 (1979).

²⁴B. Levy and G. Berthier, Int. J. Quant. Chem. 2, 307 (1968).

²⁵B. Levy, Chem. Phys. Lett. 4, 17 (1969); Int. J. Quant. Chem. 4, 297 (1970); Chem. Phys. Lett. 18, 59 (1973).

²⁶V. A. Kuprievich and O. V. Shramko, Int. J. Quant. Chem. 9, 1009 (1975); V. A. Kuprievich and V. E. Klymenko, Int. J. Quant. Chem. 10, 941 (1976).

²⁷S. Polezzo, Theor. Chim. Acta 40, 245 (1975); P. Fantucci, S. Polezzo, and M. P. Stabilini, Theor. Chim. Acta 41, 311 (1976); P. Fantucci and S. Polezzo, Theor. Chim. Acta 45, 317 (1977).

²⁸B. T. Sutcliffe, Theor. Chim. Acta 33, 201 (1974); 39, 93 (1975).

²⁹P. D. Dacre, C. J. Watts, G. R. J. Williams, and R. McWeeny, Mol. Phys. 30, 1203 (1975).

³⁰L. G. Yaffe and W. A. Goddard III, Phys. Rev. A 13, 1682 (1976).

³¹F. W. Bobrowicz and W. A. Goddard III, in *Modern Theoretical Chemistry*, edited by H. F. Schaefer, Vol. III (Plenum, New York, 1977), p. 79.

³²J. Kendrick and I. H. Hillier, Chem. Phys. Lett. 41, 283 (1976).

³³E. Dalggaard and P. Jørgensen, J. Chem. Phys. 69, 3833 (1978).

³⁴D. L. Yeager and P. Jørgensen, J. Chem. Phys. 71, 755 (1979).

³⁵A. Golebiewski, J. Hinze, and E. Yurtsever, J. Chem. Phys. 70, 1101 (1979).

³⁶See e.g., R. S. Varga, *Matrix Iterative Analysis* (Prentice-Hall, Englewood Cliffs, New Jersey, 1962), p. 56.

³⁷S. Huzinaga, Technical Report: Approximate atomic functions, Division of Theoretical Chemistry, Department of Chemistry, University of Alberta, 1965 (unpublished).