ADVANCES IN CHEMICAL PHYSICS—VOLUME LXIX

I. Prigogine and Stuart A. Rice—Editors

AB INITIO METHODS IN QUANTUM CHEMISTRY—II

Edited by

K. P. LAWLEY

Department of Chemistry
Edinburgh University

See also Volume LXVII, Ab Initio Methods in Quantum Chemistry—I

A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS

CHICHESTER · NEW YORK · BRISBANE · TORONTO · SINGAPORE

ADVANCES IN CHEMICAL PHYSICS

VOLUME LXIX

SERIES EDITORIAL BOARD

C. J. Ballahusen, Kobenhaven Universitets Fysisk–Kemiske Institut, Kemisk Laboratorium IV, Kobenhaven, Denmark
Bruce Berne, Columbia University, Department of Chemistry, New York, New York, USA
Richard B. Bernstein, Occidental Research Corporation, Irvine, California, USA
G. Careri, Istituto di Fisica 'Gugliemo Marconi', Universita delli Studi, Piazzla delle Scienze, Rome, Italy
Morrel Cohen, Exxon Research and Engineering Company, Clinton Township, Annandale, New Jersey, USA
Karl F. Freed, The James Franck Institute, the University of Chicago, Chicago Illinois, USA
Robert Gomer, The James Franck Institute, The University of Chicago, Chicago, Illinois, USA
Raymond E. Kapral, University of Toronto, 80 St. George Street, Toronto, Ontario, Canada
William Klemperer, Department of Chemistry, Harvard University, Cambridge, Massachusetts, USA
Yu L. Klimontovitch, Moscow State University, Moscow, USSR
V. Krinski, Institute of Biological Physics, USSR Academy of Sciences, Puschino, Moscow Region, USSR
M. Mandel, Chemie-Complex der Rijks-Universiteit, Wassenaarseweg, Leiden, Netherlands
Rudy Marcus, Department of Chemistry, California Institute of Technology, Pasadena, California, USA
Peter Mazur, Institute Lorentz voor Theoretische Natuurkunde, Nieuwsteeg, Leiden, Netherlands
Gregoire Nicolis, Pool de Physique, Faculté de Sciences, Université Libre de Bruxelles, Bruxelles, Belgium
A. Pacault, Centre de Recherches Paul Pascal, Domaine Universitaire, Talance, France
Yvus Pomeau, Commissariat à L'Energie Atomique, Centre d'Etudes Nucléaires de Saclay, Division de la Physique, Gif-sur-Yvette, France
A. Rahman, Argonne National Laboratory, Argonne, Illinois, USA
P. Schuster, Institut für Theoretische Chemie und Strahlenchemie, Universität Wien, Wien, Austria
I. Shavitt, Department of Chemistry, Ohio State University, Columbus, Ohio, USA
Kazuhisa Tomita, Department of Physics, Faculty of Science, Kyoto University, Kyoto, Japan

ADVANCES IN CHEMICAL PHYSICS—VOLUME LXIX

I. Prigogine and Stuart A. Rice—Editors

AB INITIO METHODS IN QUANTUM CHEMISTRY—II

Edited by

K. P. LAWLEY

Department of Chemistry
Edinburgh University

See also Volume LXVII, Ab Initio Methods in Quantum Chemistry—I

A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS

CHICHESTER · NEW YORK · BRISBANE · TORONTO · SINGAPORE

Copyright © 1987 by John Wiley & Sons Ltd.

All rights reserved.

No part of this book may be reproduced by any means, or transmitted, or translated into a machine language without the written permission of the publisher

Library of Congress Cataloging-in-Publication Data:
(Revised for vol. 2)

Ab initio methods in quantum chemistry.
(Advances in chemical physics; v. 67.)
'A Wiley-Interscience publication.'
Includes index.
1. Quantum chemistry. I. Lawley, K. P.
II. Series: Advances in chemical physics; v. 67, etc.
QD453.A27 [QD462.5] 541.2'8 86-9168

ISBN 0 471 90901 7

British Library Cataloguing in Publication Data:

Ab initio methods in quantum chemistry
part II.—(Advances in chemical physics,
ISSN 0065-2385; V.69)
1. Quantum chemistry
I. Lawley, K. P. II. Series
541.2'8 QD462

ISBN 0 471 90901 7

Printed and bound in Great Britain

CONTRIBUTORS TO VOLUME LXIX

D. L. COOPER, Department of Inorganic, Physical and Industrial Chemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

B. I. DUNLAP, Code 6129, Naval Research Laboratory, Washington, DC 20375-5000, USA

J. GERRAT, Department of Theoretical Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK

J. ODDERSHEDE, Department of Chemistry, Odense University, DK-5230 Odense M, Denmark

P. PULAY, Department of Chemistry, University of Arkansas, Fayetteville, Arkansas 72701, USA

M. RAIMONDI, Departimento di Chimica Fisica ed Electrochimica, Universita di Milano, Via Golgi 19, 20133 Milano, Italy

B. O. ROOS, Department of Theoretical Chemistry, Chemical Centre, PO Box 124, S-221 00 Lund, Sweden

D. R. SALAHUB, Département de Chimie, Université de Montréal, CP 6128, Succ. A, Montréal, Québec H3C 3J7, Canada

R. SHEPARD, Theoretical Chemistry Group, CHM, Argonne National Laboratory, Argonne, Illinois 60439, USA

F. B. VAN DUIJNEVELDT, Rijkuniversiteit Utrecht, Vakgroep Theoretische Chemie, Padualaan 8, De Uithof, Utrecht, The Netherlands

J. G. C. M. VAN DUIJNEVELDT-VAN DE RIJDT, Rijkuniversiteit Utrecht, Vakgroep Theoretische Chemie, Padualaan 8, De Uithof, Utrecht, The Netherlands

J. H. VAN LENTHE, Rijkuniversiteit Utrecht, Vakgroep Theoretische Chemie, Padualaan 8, De Uithof, Utrecht, The Netherlands

H.-J. WERNER, Institut für Physikalische and Theoretische Chemie, Johann Wolfgang Goethe-Universität Frankfurt am Main, Postfach 11 19 32, D-6000 Frankfurt, West Germany

INTRODUCTION

Few of us can any longer keep up with the flood of scientific literature, even in specialized subfields. Any attempt to do more, and be broadly educated with respect to a large domain of science, has the appearance of tilting at windmills. Yet the synthesis of ideas drawn from different subjects into new, powerful, general concepts is as valuable as eve, and the desire to remain educated persists in all scientists. This series, *Advances in Chemical Physics*, is devoted to helping the reader obtain general information about a wide variety of topics in chemical physics, which field we interpret very broadly. Our intent is to have experts present comprehensive analyses of subjects of interest and to encourage the expression of individual points of view. We hope that this approach to the presentation of an overview of a subject will both stimulate new research and serve as a personalized learning text for beginners in a field.

ILYA PRIGOGINE

STUART A. RICE

CONTENTS

MATRIX-FORMULATED DIRECT MULTICONFIGURATION SELF-CONSISTENT FIELD AND MULTICONFIGURATION REFERENCE CONFIGURATION-INTERACTION METHODS 1
H.-J. Werner

THE MULTICONFIGURATION SELF-CONSISTENT FIELD METHOD 63
R. Shepard

PROPAGATOR METHODS 201
J. Oddershede

ANALYTICAL DERIVATIVE METHODS IN QUANTUM CHEMISTRY 241
P. Pulay

SYMMETRY AND DEGENERACY IN $X\alpha$ AND DENSITY FUNCTIONAL THEORY 287
B. I. Dunlap

MODERN VALENCE BOND THEORY 319
D. L. Cooper, J. Gerratt and M. Raimondi

THE COMPLETE ACTIVE SPACE SELF-CONSISTENT FIELD METHOD AND ITS APPLICATIONS IN ELECTRONIC STRUCTURE CALCULATIONS 399
B. O. Roos

TRANSITION-METAL ATOMS AND DIMERS 447
D. R. Salahub

WEAKLY BONDED SYSTEMS 521
J. H. van Lenthe, J. G. C. M. van Duijneveldt-van de Rijdt and F. B. van Duijneveldt

AUTHOR INDEX 567

COMPOUND INDEX 583

SUBJECT INDEX 585

Ab Initio Methods in Quantum Chemistry—II
Edited by K. P. Lawley
© 1987 John Wiley & Sons Ltd.

MATRIX-FORMULATED DIRECT MULTICONFIGURATION SELF-CONSISTENT FIELD AND MULTICONFIGURATION REFERENCE CONFIGURATION-INTERACTION METHODS

HANS-JOACHIM WERNER

Institut für Physikalische und Theoretische Chemie, Johann Wolfgang Goethe-Universität Frankfurt am Main, Postfach 11 19 32, D-6000 Frankfurt, West Germany

CONTENTS

I. Introduction 2
II. Second-order Direct Multiconfiguration Self-consistent Field Theory . 5
  A. Definition of Orbitals, Density Matrices and Integral Matrices. . 5
  B. The Newton–Raphson Method and Related Optimization Procedures 7
  C. Second-order Energy Approximations 12
  D. The Variational Conditions 14
  E. Solution of the Non-linear Equations 16
  F. Optimization of Internal Orbital Rotations 20
  G. Treatment of Closed Shells 24
  H. A Direct Configuration-interaction Method for Complete Active Space Calculations 27
  I. Optimization of Energy Averages of Several States 32
III. The Internally Contracted Multiconfiguration Reference Self-consistent Electron-pairs Method 33
  A. Introduction 33
  B. The Internally Contracted Configuration Space. 38
  C. Orthogonalization and Normalization of the Configurations . . 41
  D. Coefficient Matrices and Coefficient Vectors 44
  E. The Hamilton Matrix Elements. 45
  F. Matrix Formulation of the Residual Vector 48
  G. Treatment of Closed Shells 51
  H. Evaluation of the Coupling Coefficients 54
  I. Optimization of the Contraction Coefficients 57

IV. Summary 59
Acknowledgements 59
References 60

I. INTRODUCTION

During the last 10 years, remarkable progress has been made in devising efficient procedures for accurate electronic structure calculations. Not only have the size and complexity of the problems which can be handled been extended, but also the reliability of quantum-chemical results has been considerably improved. Accurate quantum-chemical predictions of properties for small molecules have proven that theoretical calculations are often a useful complement to modern experimental work[1].

The progress in the calculation of highly correlated electronic wavefunctions is due both to the development of improved computational methods and to the rapidly increasing computing power available. In particular, the advent of vector computers has made it possible to perform much larger calculations than before in shorter times. In order to use such machines efficiently, it is essential to adjust the methods to the hardware available. Generally important is to remove all logic from the innermost loops and to perform as many simple vector or matrix operations as possible.

A central role among the available quantum-chemical tools is played by the multiconfiguration self-consistent field (MCSCF) and the multiconfiguration reference configuration-interaction (MR-CI) methods. The purpose of MCSCF calculations is to obtain electronic wavefunctions which represent the states under consideration at all investigated geometries at least qualitatively correctly. In order to obtain more accurate potential energy functions and to make reliable predictions for molecular properties, highly correlated electronic wavefunctions are necessary. It is usually impracticable to calculate such wavefunctions using the MCSCF method. Instead, one employs the MCSCF wavefunction as a zeroth-order approximation in an extended MR-CI calculation. In most MR-CI methods all single and double excitations (SD) relative to the MCSCF configurations are taken into account, and their coefficients are determined variationally. Since the length of such configuration expansions can be very large, conventional MR-CI methods[2-4] often require a configuration selection. However, during the last five years, efficient 'direct CI' methods have been developed which can handle very large configuration spaces.

The direct CI method was proposed in 1972 by Roos[5]. The idea of this method is to avoid the explicit construction and storage of the large Hamilton matrix. Instead, the eigenvectors are found iteratively. The basic operation in each iteration is to form the vector $\mathbf{g} = \mathbf{H}\cdot\mathbf{c}$ directly from the molecular integrals and the trial vector $\mathbf{c}$. The optimum algorithm to form this product depends on the structure of the wavefunction. For instance, in full CI calculations, in which all possible configurations within a given orbital basis are taken into account, it is virtually impossible to account explicitly for all different types of interactions. The necessary structure constants ('coupling coefficients') must then be calculated by a general method and stored on a formula file. This formula file is processed together with the molecular integrals in each iteration. On the other hand, if the wavefunction comprises all single and double excitations relative to a single Slater determinant (CI(SD)) only, the different configuration types and structure constants can be considered explicitly in a computer program. But even then, Roos and Siegbahn[6] distinguished more than 250 distinct cases for a closed-shell reference function. Their algorithm involved so much logic that it was rather inefficient and unsuitable for vectorization. A breakthrough for the closed-shell case was achieved by Meyer[7] in 1976 with his theory of 'self-consistent electron pairs' (SCEP). He showed that the calculation of $\mathbf{H}\cdot\mathbf{c}$ can be performed in terms of simple matrix operations, namely matrix multiplications, if the configurations are renormalized in a particular way. Particularly important in this development was that any dependence of the coupling coefficients on external orbital labels had been removed. Only very few and simple structure constants independent of the size of the basis set remained. An even more elegant formulation of the closed-shell SCEP method was recently presented by Pulay, Saebo and Meyer[8]. They were able to remove the coupling coefficients entirely and further reduce the computational effort. Because of its matrix structure, the SCEP procedure is optimally well suited for vectorization. It has been programmed and applied by Dykstra[9,10], Werner and Reinsch[11] and Ahlrichs[12].

It took a rather long time until similar matrix-formulated direct CI methods became available for more general wavefunctions. Dykstra generalized the SCEP theory for certain types of open-shell wavefunctions and for generalized valence bond (GVB) reference functions which consist only of closed-shell determinants[13,14]. In both cases, however, the important semi-internal configurations were not considered. Flesch and Meyer have developed an SCEP procedure for a spin-unrestricted Hartree–Fock (UHF) reference determinant[15]. In 1981 Chiles and Dykstra[16] presented a matrix formulation of Cicek's coupled-cluster theory. The first generalization of the SCEP method for arbitrary multiconfiguration reference wavefunctions was achieved by Werner and Reinsch[17,18] in 1981. They showed that even in this most general case the vector $\mathbf{H}\cdot\mathbf{c}$ can be obtained by performing a sequence of matrix multiplications and that all required coupling coefficients depend only on internal orbitals. In several applications the high efficiency of the method was demonstrated. Closely related matrix formulations of the direct MR-CI (SD) method were also given by Ahlrichs[19] and Saunders and van Lenthe[20].

Other multiconfiguration reference CI(SD) methods mostly used particular orthonormal spin-eigenfunction bases and employed group theoretical methods to evaluate the coupling coefficients. Most successful in this respect was the 'graphical unitary group approach' (GUGA) of Paldus[21,22] and Shavitt[23-25]. Siegbahn[26] was the first to succeed in developing a general direct MR-CI method using this technique. Other MR-CI methods were described by Buenker and Peyerimhoff[2,3], Brooks and Schaefer[27,28], Duch and Karwowski[29], Duch[30], Tavan and Schulten[31], Taylor[32], Liu and Yoshimine[33], Lischka *et al.*[34] and Saxe *et al.*[35]. Most of these methods used the fact that the external parts of the coupling coefficients are rather simple but did not eliminate them. In some of these methods, SCEP-like techniques were implemented later on (see, e.g., Ref. 36 and other articles in the same volume).

The MR-CI method usually yields most accurate results if the reference function has been fully optimized by an MCSCF procedure. In the MCSCF method not only the linear configuration coefficients but also the molecular orbitals are optimized. Owing to often strong couplings among the non-linear parameters describing changes of the molecular orbitals, early MCSCF strongly suffered with convergence difficulties. It would be beyond the scope of this chapter to review the numerous attempts to solve this problem. Considerable progress was made only quite recently with the development of second-order MCSCF methods[37-56] or approximate second-order methods[57-60]. In second-order MCSCF procedures the first and second derivatives of the energy with respect to all variational parameters, namely the orbital and configuration coefficients, are evaluated exactly. The energy is then approximated by a Taylor expansion, and the parameters are obtained by searching for a stationary point of this approximation. Close to the final solution, this method converges quadratically. Unfortunately, the radius of convergence is rather small. Therefore, many damping schemes and level-shift procedure schemes have been proposed with the aim of ensuring global convergence. This has been discussed in detail by Olsen, Yeager and Jørgensen in a previous volume of this series[53].

Another possibility to increase the radius of convergence is to include higher energy derivatives into the Taylor expansion[61,49]. The exact calculation of these derivatives is rather expensive, however. In 1980 it has been demonstrated by Werner and Meyer[42,43] that the radius of convergence can be improved considerably by treating the higher-order effects in an approximate manner. Only those terms were considered which account appropriately for the orthonormality condition of the orbitals. The extra effort to include these terms into a second-order MCSCF procedure is small. Recently, this method was further improved by Werner and Knowles[55,56], and a remarkable enhancement of convergence was achieved. Moreover, a new direct CI method devised by Knowles and Handy[62] was incorporated into the MCSCF procedure[56]. This allows one to optimize much longer configuration expansions than with previous methods.

The purpose of the present chapter is to describe in some detail the MCSCF and MCSCF-SCEP methods developed by the present author together with W. Meyer[42,43], E. A. Reinsch[17,18,63] and P. J. Knowles[55,56]. It is intended to require only little background. We hope that the rather explicit formulation will help the reader to understand various possible computational strategies, and give some insight into basic structures underlying both theories. In fact, our MCSCF and MCSCF-SCEP methods have many similarities. In both cases the integrals, the variational parameters and the coupling coefficients are ordered into matrices or vectors, and the quantities needed in each iteration are obtained by multiplying and linearly combining these matrices. Therefore, both methods are well able to be vectorized, which is important for efficient use of modern computer hardware. A reformulation of the SCEP theory in terms of non-orthogonal configurations is presented, and new techniques for efficient evaluation of the coupling coefficients are discussed. Since the emphasis of the present chapter is on the theoretical methods, only few examples for applications of our procedures are given. For a review of recent applications, the reader is referred to Ref. 64.

II. SECOND-ORDER DIRECT MULTICONFIGURATION SELF-CONSISTENT FIELD THEORY

A. Definition of Orbitals, Density Matrices and Integral Matrices

We consider a normalized N-electron wavefunction of the form

$$\Psi = \sum_I c_I \Phi_I \qquad \text{with} \qquad \langle \Phi_I | \Phi_J \rangle = \delta_{IJ} \qquad \sum_I c_I^2 = 1 \tag{1}$$

where $\{\Phi_I\}$ is a set of orthonormal configuration state functions (CSFs). Usually, the CSFs are symmetry adapted linear combinations of Slater determinants, but it is also possible to use the Slater determinants themselves as a basis. In the latter case one has to ensure in the optimization process that the wavefunction Ψ has the required symmetry. This will be discussed in more detail in Section II.H. The CSFs are constructed from the 'internal' subset of the orthonormal molecular orbitals $\{\phi_i\}$. Throughout this paper the internal orbitals will be labelled by the indices $i, j, k, \ldots$. The complementary space of external orbitals will be labelled $a, b, c \ldots$, and $r, s, t \ldots$ will denote any orbitals. The molecular orbitals (MOs) $\{\phi_r\}$ are approximated as linear combinations of atomic orbitals (AOs) or other suitable basis functions $\{\kappa_\mu\}$:

$$\phi_r = \sum_\mu X_{\mu r} \kappa_\mu \tag{2}$$

We assume that the orbitals are real and that the wavefunction is spin-restricted in the sense that each orbital can be occupied by two electrons with opposite spin. In terms of the expansion coefficients $\mathbf{X}$, the orthonormality condition of the orbitals takes the form

$$\langle r|s\rangle = (\mathbf{X}^{+}\mathbf{S}\mathbf{X})_{rs} = \delta_{rs} \tag{3}$$

where

$$S_{\mu\nu} = \langle \kappa_\mu | \kappa_\nu \rangle \tag{4}$$

is the metric of the basis $\{\kappa_\mu\}$.

The energy expectation value of the wavefunction (1) can generally be written in the form

$$E^{(0)} = \sum_{IJ} c_I c_J \left(\sum_{ij} h_{ij}\gamma_{ij}^{IJ} + \tfrac{1}{2}\sum_{ijkl} (ij|kl)\Gamma_{ij,kl}^{IJ} \right) \tag{5}$$

where h_{ij} and $(ij|kl)$ are the one-electron and two-electron integrals in the MO basis, respectively, and

$$\gamma_{ij}^{IJ} = \langle \Phi_I | E_{ij} | \Phi_J \rangle \tag{6}$$

$$\Gamma_{ij,kl}^{IJ} = \langle \Phi_I | E_{ij,kl} | \Phi_J \rangle \tag{7}$$

are coupling coefficients. The one- and two-particle excitation operators E_{ij} and $E_{ij,kl}$ in Eqs (6) and (7) are defined as follows:

$$E_{ij} = \eta_i^{\alpha+}\eta_j^{\alpha} + \eta_i^{\beta+}\eta_j^{\beta} \tag{8}$$

$$E_{ij,kl} = \eta_k^{\alpha+} E_{ij} \eta_l^{\alpha} + \eta_k^{\beta+} E_{ij} \eta_l^{\beta} \tag{9}$$

Here η_i^{α} and $\eta_i^{\beta+}$ are the usual annihilation and creation operators for electrons with α and β spin, respectively. In order to satisfy the Pauli exclusion principle, they must obey the anticommutation relations

$$[\eta_i^{\rho}, \eta_j^{\sigma}]_+ = 0 \tag{10}$$

$$[\eta_i^{\rho+}, \eta_j^{\sigma+}]_+ = 0 \tag{11}$$

$$[\eta_i^{\rho+}, \eta_j^{\sigma}]_+ = \delta_{ij}\delta_{\rho\sigma} \qquad \text{with} \qquad \rho,\sigma = \{\alpha,\beta\} \tag{12}$$

The coupling coefficients γ_{ij}^{IJ} and $\Gamma_{ij,kl}^{IJ}$ depend only on the formal structure of the CSFs $\{\Phi_I\}$ but not on the particular form of the orbitals involved. They can, therefore, be calculated once and stored on a formula tape. However, as will be discussed later, it is in certain cases advantageous to re-evaluate them each time they are needed. From the coupling coefficients and CI coefficients, the first- and second-order density matrices can be obtained:

$$(\mathbf{D})_{rs} = \sum_{IJ} c_I c_J \gamma_{rs}^{IJ} \tag{13}$$

$$\Gamma_{rs,tu} = \sum_{IJ} c_I c_J \Gamma_{rs,tu}^{IJ} \tag{14}$$

For convenience in later expressions, we define the symmetrized density matrices

$$(\mathbf{P}^{kl})_{rs} = \tfrac{1}{2}(\Gamma_{rs,kl} + \Gamma_{sr,kl}) \tag{15}$$

$$(\mathbf{Q}^{kl})_{rs} = \tfrac{1}{2}(\Gamma_{rk,sl} + \Gamma_{kr,sl}) \tag{16}$$

Note that for MCSCF wavefunctions these matrices have non-vanishing elements only in the internal–internal block. According to the above definitions the following symmetry relations hold for real orbitals:

$$(\mathbf{D})_{ij} = (\mathbf{D})_{ji} \tag{17}$$

$$(\mathbf{P}^{kl})_{ij} = (\mathbf{P}^{kl})_{ji} = (\mathbf{P}^{ij})_{kl} \tag{18}$$

$$(\mathbf{Q}^{kl})_{ij} = (\mathbf{Q}^{lk})_{ji} = (\mathbf{Q}^{ij})_{kl} \tag{19}$$

We further define Coulomb and exchange matrices

$$(\mathbf{J}^{kl})_{rs} = (rs|kl) \tag{20}$$

$$(\mathbf{K}^{kl})_{rs} = (rk|ls) \tag{21}$$

which are ordered subsets of the two-electron integrals. This ordering of the integrals with at most two external orbitals is essential for the matrix formulation of both MCSCF and MCSCF-CI methods. As indicated by the parentheses in Eqs (18)–(21), superscripts denote different matrices, and subscripts their elements. This convention will be followed throughout this chapter and the parentheses will often be omitted. The energy expectation value can now be written in the simple form

$$E^{(0)} = \mathrm{tr}(\mathbf{hD}) + \tfrac{1}{2}\sum_{kl} \mathrm{tr}(\mathbf{J}^{kl}\mathbf{P}^{lk}) \tag{22}$$

where tr(**A**) denotes the trace of the matrix **A**.

B. The Newton–Raphson Method and Related Optimization Procedures

The purpose of the MCSCF method is to minimize the energy expectation value (Eq. (5)) with respect to the CI coefficients $\{c_I\}$ and the molecular-orbital coefficients $X_{\mu i}$ with the auxiliary orthonormality condition in Eq. (3). Since the energy expectation value is a fourth-order function of the orbitals, its direct minimization is impracticable. It is, therefore, necessary to employ an iterative procedure and minimize in each iteration an approximate energy functional. Provided this functional is a reasonably good approximation to the true energy as a function of the changes in the orbitals and CI coefficients, its minimization will yield improved orbitals and CI coefficients. These are used as a starting guess in the next iteration. If the approximate energy functional is accurate to second order in the changes of the orbitals and CI coefficients, the optimization will be quadratically convergent, i.e. when approaching the final solution the change of the energy in each iteration decreases quadratically. Far from the solution, however, the approximate second-order functional may not be a reasonable approximation. Special precautions are then necessary to ensure convergence. In fact, extensive work of several groups has been performed during the last few years in order to find stable and efficient algorithms which improve or even guarantee convergence in the non-local region[39–41,46,48,50–53].

The common approach in most methods is to describe the orbital changes by a unitary (orthogonal) transformation of the form[65,66]

$$|\tilde{i}\rangle = \sum_r |r\rangle U_{ri} \tag{23}$$

i.e.

$$\tilde{\mathbf{X}} = \mathbf{XU} \tag{24}$$

with

$$\mathbf{U} = \exp(\mathbf{R}) = \mathbf{1} + \mathbf{R} + \tfrac{1}{2}\mathbf{RR} + \cdots \tag{25}$$

where $\mathbf{R} = -\mathbf{R}^\dagger$ is an antisymmetric matrix. Since $\mathbf{U}$ remains unitary for any choice of this matrix, the elements $\{R_{ri}, r > i\}$ form a set of independent variational parameters. Some of the parameters R_{ij} may be redundant, i.e. they do not influence the energy to first order. This happens if the orbitals $|i\rangle$ and $|j\rangle$ have the same occupation number in all CSFs. An R_{ij} is also redundant if the same first-order energy change can be achieved by a variation of the CI coefficients. Since redundant variables do influence the energy in higher order, they must be set to zero in order to avoid convergence difficulties. The redundant variables can be determined automatically as described in Ref. 67.

In the Newton–Raphson method the energy is expanded up to second order in the variables R_{ri} and the changes of the CI coefficients $\{\Delta c_I\}$. Collecting these parameters into a vector $\mathbf{x}$, the stationary condition for the energy approximation

$$E^{(2)}(\mathbf{x}) = E^{(0)} + \mathbf{g}^\dagger\mathbf{x} + \tfrac{1}{2}\mathbf{x}^\dagger\mathbf{Hx} \tag{26}$$

takes the form of a system of inhomogeneous linear equations

$$\mathbf{Hx} + \mathbf{g} = 0 \tag{27}$$

where

$$g_i = (\partial E/\partial x_i)_{\mathbf{x}=\mathbf{0}} \tag{28}$$

is the energy gradient at the expansion point, and

$$H_{ij} = (\partial^2 E/\partial x_i \partial x_j)_{\mathbf{x}=\mathbf{0}} \tag{29}$$

is the Hessian matrix of second energy derivatives. Explicit expressions for these derivatives will be given in Section II.C. The solution of Eq. (27) yields the parameters $\mathbf{x}$, which are used to calculate new orbitals and a new CI vector. Experience has shown the radius of convergence of this method to be rather small. Far from the solution the Hessian often has many negative or very small eigenvalues, and convergence can then be achieved only by introducing a level shift which makes the Hessian positive definite and the step vector $\mathbf{x}$ sufficiently small:

$$(\mathbf{H} + \kappa\mathbf{1})\mathbf{x} + \mathbf{g} = 0 \tag{30}$$

If the level shift is chosen to be

$$\kappa = -\lambda\varepsilon \tag{31}$$

with

$$\varepsilon = \lambda\mathbf{g}^{\dagger}\mathbf{x} \tag{32}$$

the linear equations are transformed into an eigenvalue equation of the form:

$$\begin{pmatrix} -\varepsilon & \mathbf{g}^{\dagger} \\ \mathbf{g} & \mathbf{H}/\lambda - \varepsilon \end{pmatrix}\begin{pmatrix} 1/\lambda \\ \mathbf{x} \end{pmatrix} = \mathbf{0} \tag{33}$$

For $\lambda = 1$ this is known as the augmented Hessian (AH) method. It was first proposed by Lengsfield[39], and used with various modifications by several authors[44,46,51,52]. It can easily be proved that $\mathbf{H} - \varepsilon\mathbf{1}$ is always positive definite if ε is the eigenvalue obtained by solving Eq. (33). It can also be shown that the AH method is quadratically convergent[46]. A value $\lambda > 1$ has the effect of further reducing the step length $|\mathbf{x}|$. In fact, as shown by Fletcher[68] and

TABLE I
Convergence behaviour of step-restricted augmented Hessian calculations (Fletcher optimization).

Iter.	Energy difference		
	N_2[a]	CO[b]	CO[c]
1	−0.019350483	−0.016819128	−0.005805628
2	−0.016995391	−0.017490633	−0.015098631
3	−0.016661071	−0.013430376	−0.027426813
4	−0.002740679	−0.015795238	−0.026688321
5	−0.000038990	−0.003660917	−0.009411258
6	−0.000000013	−0.001231696	−0.018214694
7	−0.000000000	−0.000050933	−0.016727086
8		−0.000001714	−0.001016208
9		−0.000000003	−0.000039643
10		−0.000000000	−0.000000210
11			−0.000000000

[a] N_2 molecule; for details see Ref. 51; improved virtual orbitals (IVOs) were taken as starting guess.
[b] CO molecule; for details see Ref. 51; IVOs were used as starting guess.
[c] CO molecule; basis set and configurations as in footnote *b*, but canonical SCF orbitals used as starting guess; results from Ref. 55.

discussed in the context of the MCSCF problem by Jørgensen *et al.*[51,52,54], Eq. (33) can be derived by minimizing the second-order energy approximation (Eq. (26)) with the auxiliary condition $|\mathbf{x}| < s$. If the maximum step length s is updated automatically according to a particular scheme after each iteration, convergence can be guaranteed. This does not mean, however, that convergence is achieved with a small number of iterations as desired for a second-order scheme. Test calculations published by Jørgensen *et al.* have shown that typically 6–10 iterations are necessary unless a very good starting guess from a nearby geometry is available (cf. Table I). We have similar experiences using this method. Of these iterations, only the last two or three are in the local region and show quadratic convergence behaviour.

The origin of these difficulties is the orthonormality condition in Eq. (3). This causes the true energy to be periodic in individual orbital rotations. If only a single rotation between the orbitals $|i\rangle$ and $|j\rangle$ is considered, the unitary

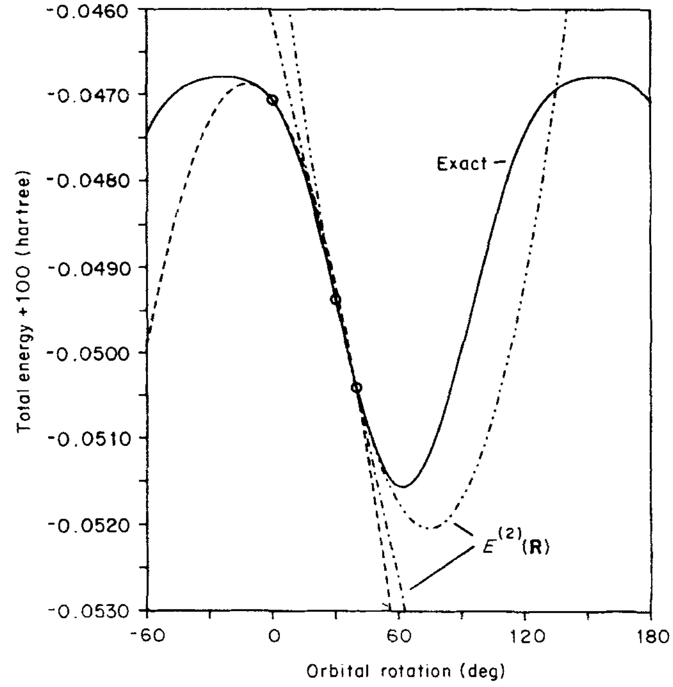

Fig. 1. The dependence of the exact energy and the second-order energy approximation $E^{(2)}(\mathbf{R})$ on the rotation 4σ–5σ for a three-configuration MCSCF calculation for the HF molecule. The configurations were: $1\sigma^2 2\sigma^2 3\sigma^2 1\pi^4$, $1\sigma^2 2\sigma^2 4\sigma^2 1\pi^4$ and $1\sigma^2 3\sigma^2 4\sigma^2 1\pi^4$. For other details see Ref. 55. The expansion point at zero degrees corresponds to canonical SCF orbitals. $E^{(2)}(\mathbf{R})$ is also shown for two other expansion points. In these cases all orbitals except 4σ and 5σ were canonical SCF orbitals.

matrix **U** can be written in the form

$$\mathbf{U} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \tag{34}$$

where the rotation angle α equals R_{ij}. Clearly, as illustrated in Fig. 1, the second-order energy approximation does not describe this periodicity. Its minimization, therefore, predicts steps that are either too large or even of the wrong sign. Fig. 2 shows the effect of level shifts which restrict the step size to certain values. It is obvious that convergence will depend sensitively on the choice of the step size. Of course, Figs 1 and 2 are very idealized examples. In cases with many orbital rotations which influence each other, the situation is much more complicated, and a single level-shift parameter cannot be expected to be optimal for all orbital rotations.

From the above considerations it appears to be necessary to account more accurately, in the energy approximation, for the orthonormality of the orbitals. A straightforward extension of the Newton–Raphson method would be to expand the energy up to third or even higher order in $\mathbf{R}^{49,53,61}$.

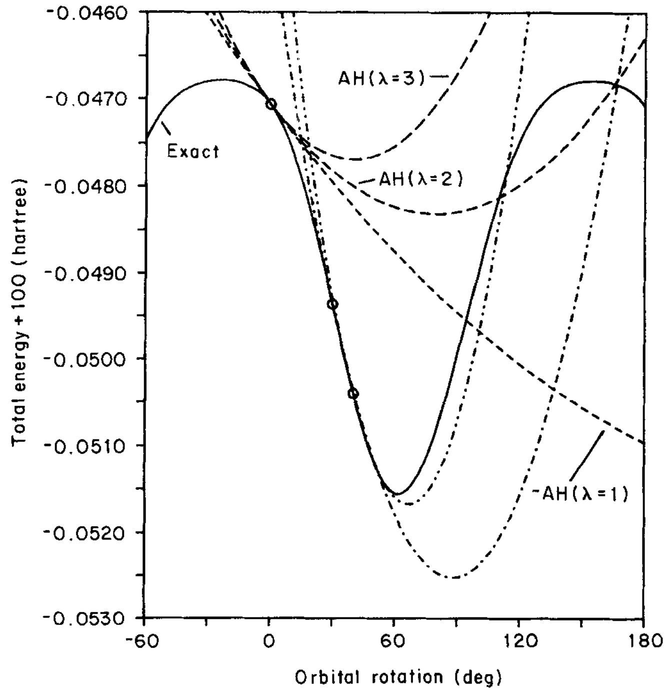

Fig. 2. Second-order energy approximations employed in step-restricted augmented Hessian calculations for the same model and the same expansion points as in Fig. 1. For the expansions at zero degrees, various level-shift parameters have been used.

However, this would be only a partial improvement, since the true orbital and energy changes are of infinite order in **R**. Furthermore, each iteration would require a considerably more expensive integral transformation than a second-order scheme. Fortunately, it is possible to account for the orthonormality of the orbitals appropriately in a rather simple way already in a second-order method. This will be explained in the next sections.

C. Second-order Energy Approximations

The energy expectation value (Eq. (5)) is a function of the CI coefficients $\{c_I\}$ and the orbital changes

$$|\Delta i\rangle = |\tilde{i}\rangle - |i\rangle = \sum_r |r\rangle T_{ri} \tag{35}$$

where

$$\mathbf{T} = \mathbf{U} - \mathbf{1} = \mathbf{R} + \tfrac{1}{2}\mathbf{R}\mathbf{R} + \cdots \tag{36}$$

The dependence on the CI coefficients enters via the density matrices **D** and $\mathbf{P}^{kl}$ and will be considered explicitly later on. The exact energy is a fourth-order function of the orbital changes, i.e. $E = E^{(4)}(\mathbf{T})$, and of infinite order in **R**. If the energy expansion is truncated to second order in the orbital changes, one obtains

$$\begin{aligned} E^{(2)} = E^{(0)} &+ 2\sum_{ij} \langle \Delta i|h|j\rangle D_{ij} + \sum_{ij} \langle \Delta i|h|\Delta j\rangle D_{ij} \\ &+ \sum_{ijkl} [2(\Delta ij|kl)P_{ij}^{kl} + (\Delta i\Delta j|kl)P_{ij}^{kl} + 2(\Delta ik|l\Delta j)Q_{ij}^{kl}] \end{aligned} \tag{37}$$

where the symmetry relations in Eqs (17)–(19) have been used to sum equivalent terms. In terms of **U** and **T**, Eq. (37) takes the form

$$\begin{aligned} E^{(2)}(\mathbf{T}) = E^{(0)} &+ 2\,\mathrm{tr}(\mathbf{T}^\dagger\mathbf{h}\mathbf{D}) + \mathrm{tr}(\mathbf{T}^\dagger\mathbf{h}\mathbf{T}\mathbf{D}) \\ &+ \sum_{kl} [2\,\mathrm{tr}(\mathbf{T}^\dagger\mathbf{J}^{kl}\mathbf{P}^{lk}) + \mathrm{tr}(\mathbf{T}^\dagger\mathbf{J}^{kl}\mathbf{T}\mathbf{P}^{lk}) + 2\,\mathrm{tr}(\mathbf{T}^\dagger\mathbf{K}^{kl}\mathbf{T}\mathbf{Q}^{lk})] \end{aligned} \tag{38}$$

Defining the matrices

$$\mathbf{A} = \mathbf{h}\mathbf{D} + \sum_{kl} \mathbf{J}^{kl}\mathbf{P}^{lk} \tag{39}$$

$$\begin{aligned} \mathbf{B} &= \mathbf{A} + \mathbf{h}\mathbf{T}\mathbf{D} + \sum_{kl} (\mathbf{J}^{kl}\mathbf{T}\mathbf{P}^{lk} + 2\mathbf{K}^{kl}\mathbf{T}\mathbf{Q}^{lk}) \\ &= \mathbf{h}\mathbf{U}\mathbf{D} + \sum_{kl} (\mathbf{J}^{kl}\mathbf{U}\mathbf{P}^{lk} + 2\mathbf{K}^{kl}\mathbf{T}\mathbf{Q}^{lk}) \end{aligned} \tag{40}$$

and

$$\mathbf{G}^{ij} = \mathbf{h}D_{ij} + \sum_{kl} (\mathbf{J}^{kl}P_{ij}^{kl} + 2\mathbf{K}^{kl}Q_{ij}^{kl}) \tag{41}$$

the second-order energy can also be written in the more compact forms:

$$\begin{aligned} E^{(2)}(\mathbf{T}) &= E^{(0)} + 2\,\mathrm{tr}(\mathbf{T}^{\dagger}\mathbf{A}) + \sum_{ij}(\mathbf{T}^{\dagger}\mathbf{G}^{ij}\mathbf{T})_{ij} \\ &= E^{(0)} + \mathrm{tr}[\mathbf{T}^{\dagger}(\mathbf{A}+\mathbf{B})] \end{aligned} \tag{42}$$

Note that, owing to the sparsity of the density matrices $\mathbf{D}$, $\mathbf{P}^{kl}$ and $\mathbf{Q}^{kl}$, all elements A_{ra} and B_{ra} ($|a\rangle$ external) vanish. Therefore, in a computer program only the rectangular blocks A_{ri} and B_{ri} have to be computed and stored. For the sake of compact expressions, however, it is advantageous to deal formally with the full square matrices.

As outlined in Section II.B, in the Newton–Raphson (NR) method the energy approximation in Eq. (42) is truncated to second order in $\mathbf{R}$. The explicit form of Eq. (26) for fixed CI coefficients is therefore

$$E^{(2)}(\mathbf{R}) = E^{(0)} + 2\,\mathrm{tr}(\mathbf{R}^{\dagger}\mathbf{A}) + \mathrm{tr}(\mathbf{RRA}) + \sum_{ij}(\mathbf{R}^{\dagger}\mathbf{G}^{ij}\mathbf{R})_{ij} \tag{43}$$

In contrast to this approximation, Eq. (42) contains terms up to infinite order in $\mathbf{R}$. These additional terms account appropriately for the orthonormality of the orbitals. In fact, as shown in Fig. 3 for the same model calculation as in Figs

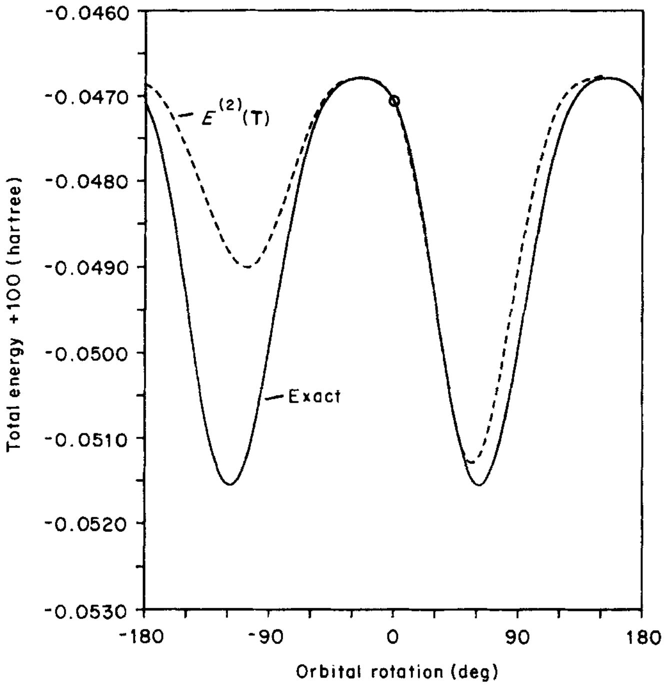

Fig. 3. The dependence of the second-order energy approximation $E^{(2)}(\mathbf{T})$ on the rotation 4σ–5σ for the HF molecule as in Fig. 1.

1 and 2, $E^{(2)}(\mathbf{T})$ is in close agreement with the true energy over a large range of rotation angles and predicts minima at nearly the correct angles. Thus, a much larger radius of convergence can be expected if $E^{(2)}(\mathbf{T})$ rather than $E^{(2)}(\mathbf{R})$ is used as approximate energy functional. It will be demonstrated also that the rate of convergence in the non-local region is much faster* than for a step-restricted augmented Hessian method.

We will now investigate how the second-order energy $E^{(2)}(\mathbf{T})$ at a particular point $\mathbf{T} = \mathbf{T}(\mathbf{R})$ changes if $\mathbf{T}$ undergoes a small variation. Such a change can be described by multiplying $\mathbf{U}$ with a second unitary transformation $\mathbf{U}(\Delta\mathbf{R})$:

$$\begin{aligned} \mathbf{U}(\mathbf{R}, \Delta\mathbf{R}) &= \mathbf{U}(\mathbf{R})\mathbf{U}(\Delta\mathbf{R}) \\ &= \mathbf{U} + \mathbf{U}(\Delta\mathbf{R} + \tfrac{1}{2}\Delta\mathbf{R}\Delta\mathbf{R} + \cdots) \end{aligned} \tag{44}$$

The antisymmetric matrix $\Delta\mathbf{R} = -\Delta\mathbf{R}^{\dagger}$ defines the change of $\mathbf{U}$. Note that $\mathbf{U}(\mathbf{R}, \Delta\mathbf{R}) \neq \mathbf{U}(\mathbf{R} + \Delta\mathbf{R})$ since $\mathbf{R}$ and $\Delta\mathbf{R}$ do not commute. Inserting this into Eq. (38) yields, up to second order in $\Delta\mathbf{R}$,

$$\begin{aligned} E^{(2)}(\mathbf{T}, \Delta\mathbf{R}) &= E^{(2)}(\mathbf{T}) + 2\,\mathrm{tr}(\Delta\mathbf{R}^{\dagger}\tilde{\mathbf{A}}) + \mathrm{tr}(\Delta\mathbf{R}\Delta\mathbf{R}\tilde{\mathbf{A}}) + \sum_{ij}(\Delta\mathbf{R}^{\dagger}\tilde{\mathbf{G}}^{ij}\Delta\mathbf{R})_{ij} \\ &= E^{(2)}(\mathbf{T}) + \mathrm{tr}[\Delta\mathbf{R}^{\dagger}(\tilde{\mathbf{A}} + \tilde{\mathbf{B}})] + \mathrm{tr}(\Delta\mathbf{R}\Delta\mathbf{R}\tilde{\mathbf{A}}) \end{aligned} \tag{45}$$

where

$$\tilde{\mathbf{A}} = \mathbf{U}^{\dagger}\mathbf{B} \tag{46}$$

$$\tilde{\mathbf{B}} = \tilde{\mathbf{A}} + \mathbf{U}^{\dagger}\left(\mathbf{h}\mathbf{U}\Delta\mathbf{R}\mathbf{D} + \sum_{kl}(\mathbf{J}^{kl}\mathbf{U}\Delta\mathbf{R}\mathbf{P}^{lk} + 2\mathbf{K}^{kl}\mathbf{U}\Delta\mathbf{R}\mathbf{Q}^{lk})\right) \tag{47}$$

$$\tilde{\mathbf{G}}^{ij} = \mathbf{U}^{\dagger}\mathbf{G}^{ij}\mathbf{U} \tag{48}$$

One should note the similarity of Eqs (40) and (47). For $\mathbf{U} = \mathbf{1}$ and $\mathbf{T} = \mathbf{0}$ we have $\tilde{\mathbf{A}} = \mathbf{A}$ and $\tilde{\mathbf{G}}^{ij} = \mathbf{G}^{ij}$. Eq. (45) then reduces to Eq. (43) (Newton–Raphson approximation).

D. The Variational Conditions

The energy expectation value $E = E^{(4)}(\mathbf{T})$ in Eq. (5) has a stationary point if the first derivatives at the expansion point $\mathbf{T} = \mathbf{0}$ with respect to all R_{ri} vanish, i.e. if

$$(\partial E/\partial R_{ri})_{\mathbf{R}=\mathbf{0}} = 2(\mathbf{A} - \mathbf{A}^{\dagger})_{ri} = 0 \qquad \text{for all } r > i \tag{49}$$

The stationary point is a minimum if the Hessian matrix (Eq. (29)) is positive definite. Furthermore, the CI coefficients must satisfy the eigenvalue equation

$$\sum_{J}(H_{IJ} - E\delta_{IJ})c_J = 0 \qquad \text{for all } I \tag{50}$$

---

*This may not be true for rotations between strongly occupied orbitals, cf. Section II. F.

This implies that for the electronic ground state E is the lowest eigenvalue of the Hamilton matrix

$$\begin{aligned} H_{IJ} &= \langle \Phi_I | H | \Phi_J \rangle \\ &= \sum_{ij} h_{ij} \gamma_{ij}^{IJ} + \tfrac{1}{2} \sum_{ijkl} (ij|kl) \Gamma_{ij,kl}^{IJ} \end{aligned} \tag{51}$$

If Eqs (49) and (50) are satisfied simultaneously, convergence of the MCSCF procedure is reached.

Similarly, the energy approximation $E^{(2)}(\mathbf{T})$ has a stationary point with respect to variations of $\mathbf{T}$ if the first derivatives

$$\left(\frac{\partial E^{(2)}(\mathbf{T}, \Delta\mathbf{R})}{\partial \Delta R_{ri}} \right)_{\Delta\mathbf{R}=\mathbf{0}} = 2(\tilde{\mathbf{A}} - \tilde{\mathbf{A}}^{\dagger})_{ri} \tag{52}$$

vanish for all $r > i$. These conditions are summarized in the matrix equation[42,43,55]

$$\mathbf{U}^{\dagger}\mathbf{B} - \mathbf{B}^{\dagger}\mathbf{U} = \mathbf{0} \tag{53}$$

The stationary point is a minimum if the matrix of second derivatives

$$\left(\frac{\partial^2 E^{(2)}(\mathbf{T}, \Delta\mathbf{R})}{\partial \Delta R_{ri} \partial \Delta R_{sj}} \right)_{\Delta\mathbf{R}=\mathbf{0}} = (1 - \tau_{ri})(1 - \tau_{sj})(2\tilde{\mathbf{G}}^{ij} - \delta_{ij}(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^{\dagger}))_{rs} \tag{54}$$

is positive definite. The operator τ_{ri} in Eq. (54) permutes the indices r and i. For $\mathbf{T} = \mathbf{0}$ Eqs (52) and (54) reduce to the explicit formulae for the energy derivatives used in the NR or AH methods (Eqs (28) and (29)).

In order to minimize the second-order energy approximation $E^{(2)}(\mathbf{T}, \mathbf{c})$ with respect to the CI coefficients, it can be written in the form

$$E^{(2)}(\mathbf{T}, \mathbf{c}) = \mathbf{c}^{\dagger}\mathbf{H}^{(2)}\mathbf{c}/\mathbf{c}^{\dagger}\mathbf{c} \tag{55}$$

where the second-order Hamiltonian $\mathbf{H}^{(2)}$ is defined as

$$H_{IJ}^{(2)} = \sum_{ij} (\mathbf{U}^{\dagger}\mathbf{h}\mathbf{U})_{ij} \gamma_{ij}^{IJ} + \tfrac{1}{2} \sum_{ijkl} (ij|kl)^{(2)} \Gamma_{ij,kl}^{IJ} \tag{56}$$

The integrals $(ij|kl)^{(2)}$ are the second-order approximations to the exact two-electron integrals as a function of $\mathbf{T}$:

$$\begin{aligned} (ij|kl)^{(2)} = &-(ij|kl) + (\mathbf{U}^{\dagger}\mathbf{J}^{kl}\mathbf{U})_{ij} + (\mathbf{U}^{\dagger}\mathbf{J}^{ij}\mathbf{U})_{kl} \\ &+ (1 + \tau_{ij})(1 + \tau_{kl})(\mathbf{T}^{\dagger}\mathbf{K}^{ik}\mathbf{T})_{jl} \end{aligned} \tag{57}$$

According to this definition the second-order energy expressions in Eqs (38) and (55) are identical for a given set $\{\mathbf{c}, \mathbf{T}\}$. The minimization of the energy 'expectation value' (Eq. (55)) with respect to the $\{c_I\}$ yields the eigenvalue equation

$$(\mathbf{H}^{(2)} - E^{(2)}\mathbf{1})\mathbf{c} = 0 \tag{58}$$

The minimum of $E^{(2)}(\mathbf{T}, \mathbf{c})$ with respect to $\mathbf{T}$ and $\mathbf{c}$ is reached if the coupled non-linear equations (53) and (58) are satisfied simultaneously with the same $\mathbf{T}$ and $\mathbf{c}$. In this case the energy eigenvalue $E^{(2)}$ in Eq. (58) becomes identical with the expectation values in Eqs (38) and (55).

E. Solution of the Non-linear Equations

In order to make the optimization procedure outlined in Sections II.C and II.D practicable, a stable algorithm to solve the coupled non-linear equations (53) and (58) is necessary. The method should avoid the explicit construction and storage of large Hessian or Hamilton matrices in order to be flexible with respect to the number of orbitals and configuration state functions. Hence, for the optimization of the CI coefficients it is advantageous to employ a 'direct CI' procedure. In a direct CI method the desired eigenvector is obtained iteratively. In each iteration the 'residual vector' $\mathbf{y} = (\mathbf{H} - E\mathbf{1})\mathbf{c}$ is calculated directly from the one- and two-electron integrals, a trial vector $\mathbf{c}$ and the coupling coefficients. The residual vector is then used to improve the trial vector $\mathbf{c}$. Very similar techniques can be employed to solve iteratively large systems of linear equations or the non-linear equations (53) ('direct MCSCF'). The iterations needed to solve Eqs (53) and (58) are called 'micro-iterations'. After convergence of the micro-iterations, the final matrix $\mathbf{U}$ is used to transform the orbital coefficients according to Eq. (24). Then a new set of operators $\mathbf{J}^{kl}$, $\mathbf{K}^{kl}$ is evaluated. Efficient algorithms for this partial four-index transformation have been described by several authors[20,42,57]. The calculation of these operators and a variational energy initializes the next 'macro-iteration'.

In complete active space self-consistent field (CASSCF) calculations[57,59,69] with long configuration expansions the most expensive part is often the optimization of the CI coefficients. It is, therefore, particularly important to minimize the number of CI iterations. In conventional direct second-order MCSCF procedures[44,52,70], the CI coefficients are updated together with the orbital parameters in each micro-iteration. Since the optimization requires typically 100–150 micro-iterations, such calculations with many configurations can be rather expensive. A possible remedy to this problem is to decouple the orbital and CI optimizations[59], but this causes the loss of quadratic convergence. The following method allows one to update the CI coefficients much fewer times than the orbital parameters. This saves considerable time without loss of the quadratic convergence behaviour.

In order to minimize the second-order energy approximation $E^{(2)}(\mathbf{T})$ for fixed CI coefficients a step-restricted augmented Hessian method as outlined in Section II.B (Eqs (30)–(33)) is used. While in other MCSCF methods this technique is employed to minimize the exact energy, it is used here to minimize an approximate energy functional. The parameter vector $\mathbf{x}$ is made up of the non-redundant elements ΔR_{ri} $(r > i)$. The successive expansion points are defined by the matrix $\mathbf{T} = \mathbf{U} - \mathbf{1}$, which is updated according to Eq. (44) each time the eigenvalue equation (33) has been solved using Davidson's technique[71]. In each micro-iteration, the residual vector

$$\mathbf{y} = \mathbf{g} + (\mathbf{H} - \lambda\varepsilon)\mathbf{x} \tag{59}$$

has to be evaluated, where the elements of the gradient **g** and the Hessian **H** are given in Eqs (52) and (54), respectively. The damping parameter λ is determined automatically as the Davidson iteration proceeds such that the step length $|\mathbf{x}|$ remains smaller than a prescribed threshold (e.g. 0.5). The explicit form of the residual vector is obtained by deriving Eq. (45) with respect to all ΔR_{ri}. It can be written in matrix form as

$$\begin{aligned} \mathbf{Y} &= 2(\tilde{\mathbf{B}} - \tilde{\mathbf{B}}^\dagger) - (\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\dagger)\Delta\mathbf{R} + \Delta\mathbf{R}(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\dagger) - \lambda\varepsilon\Delta\mathbf{R} \\ \text{with} \quad & \varepsilon = 2\,\mathrm{tr}(\Delta\mathbf{R}^\dagger\tilde{\mathbf{A}}) \end{aligned} \tag{60}$$

The matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ have been defined in Eqs (46) and (47). In the Davidson procedure, the matrix $\Delta\mathbf{R}$ and the residual **Y** are obtained as linear combinations

$$\Delta\mathbf{R} = \sum_m \alpha_m \mathbf{S}^m \tag{61}$$

$$\mathbf{Y} = \sum_m \alpha_m \mathbf{Y}^m \tag{62}$$

where $\mathbf{S}^m$ are orthonormalized expansion vectors.* The $\mathbf{Y}^m$ are calculated according to Eq. (60) with the $\mathbf{S}^m$ instead of $\Delta\mathbf{R}$. The optimum parameters α_m are determined by solving a small eigenvalue problem[71]. A new expansion vector is then obtained as

$$S_{ri}^{m+1} = -\,Y_{ri}/(D_{ri} - \varepsilon) \tag{63}$$

and subsequently orthonormalized to all previous $\mathbf{S}^m$. The diagonal elements D_{ri} of the Hessian matrix (Eq. (54)) used in the update formula (63) are given by

$$D_{ai} = [2\mathbf{U}^\dagger\mathbf{G}^{ii}\mathbf{U} - (\mathbf{U}^\dagger\mathbf{B} + \mathbf{B}^\dagger\mathbf{U})]_{a\sigma} \tag{64}$$

$$\begin{aligned} D_{ji} = {} & 2[(\mathbf{U}^\dagger\mathbf{G}^{ii}\mathbf{U})_{jj} + (\mathbf{U}^\dagger\mathbf{G}^{jj}\mathbf{U})_{ii} - 2(\mathbf{U}^\dagger\mathbf{G}^{ji}\mathbf{U})_{ij}] \\ & - (\mathbf{U}^\dagger\mathbf{B} + \mathbf{B}^\dagger\mathbf{U})_{ii} - (\mathbf{U}^\dagger\mathbf{B} + \mathbf{B}^\dagger\mathbf{U})_{jj} \end{aligned} \tag{65}$$

Usually it is sufficient to calculate these elements only once per macro-iteration with $\mathbf{U} = \mathbf{1}$, such that the operators $\mathbf{G}^{ij}$ need not be transformed each time **U** is updated. If the Davidson procedure has converged (i.e. **Y** is smaller than a certain threshold) the unitary matrix **U** is updated according to Eq. (44), and a new matrix **B** is calculated. The process is repeated until Eq. (53) is satisfied to the desired accuracy. One should note that no expensive integral transformation is necessary after updating $\mathbf{U}$.

---

* In the orthonormalization process the non-redundant elements of the matrices $\mathbf{S}^m$ form a vector.

After an update of $\mathbf{U}$ it is possible to perform a direct CI step in order to improve the CI vector and the density matrices. In this case the calculation of $\mathbf{B}$ is done in two steps. First, the one-index transformations

$$(\tilde{\mathbf{h}})_{rj} = (\mathbf{h}\mathbf{U})_{rj} \tag{66}$$

$$(\tilde{\mathbf{J}}^{kl})_{rj} = (\mathbf{J}^{kl}\mathbf{U})_{rj} \tag{67}$$

$$(\tilde{\mathbf{K}}^{kl})_{rj} = (\mathbf{K}^{kl}\mathbf{T})_{rj} \tag{68}$$

are performed, and, at the same time, the second-order integrals $(ij|kl)^{(2)}$ are evaluated by performing the second half transformations

$$(\mathbf{U}^{\dagger}\mathbf{h}\mathbf{U})_{ij} = (\mathbf{U}^{\dagger}\tilde{\mathbf{h}})_{ij} \tag{69}$$

$$(ij|kl)^{(2)} \leftarrow (\mathbf{U}^{\dagger}\tilde{\mathbf{J}}^{kl})_{ij} \tag{70}$$

$$(ik|jl)^{(2)} \leftarrow (\mathbf{T}^{\dagger}\tilde{\mathbf{K}}^{kl})_{ij} \tag{71}$$

Since only the internal blocks of these transformed matrices are needed, the latter step is much cheaper than the first half of the transformation (Eqs (66)–(68)). The half-transformed operators $\tilde{\mathbf{h}}$, $\tilde{\mathbf{J}}^{kl}$ and $\tilde{\mathbf{K}}^{kl}$ are stored on disc, while the integrals $(ij|kl)^{(2)}$ are kept in high-speed memory and employed in the subsequent direct CI step. The direct CI procedure will be described in Section II.H. The improved CI vector is used to evaluate new density matrices $\mathbf{D}$, $\mathbf{P}^{kl}$ and $\mathbf{Q}^{kl}$. Finally, the new matrix $\mathbf{B}$ is calculated according to

$$\mathbf{B} = \tilde{\mathbf{h}}\mathbf{D} + \sum_{kl}(\tilde{\mathbf{J}}^{kl}\mathbf{P}^{lk} + 2\tilde{\mathbf{K}}^{kl}\mathbf{Q}^{lk}) \tag{72}$$

As compared with other direct second-order MCSCF methods[44,52,70], the above procedure has the following advantages: (i) A CI step is only performed if the transformation matrix $\mathbf{U}$ has converged to a sufficiently stable value as indicated by a small step size $\Delta\mathbf{R}$ in the previous augmented Hessian iteration. This avoids unnecessary oscillations of the CI coefficients and minimizes the number of CI steps. (ii) Each update of the CI vector requires the cost of only two direct CI iterations, one for the calculation of the residual vector and one for the evaluation of the density matrices. If the orbitals and CI coefficients are optimized by a coupled Newton–Raphson or augmented Hessian procedure, each CI update should be about three times as expensive as a simple direct CI step. In practice, even a factor of 5 has been reported[70]. (iii) Since the density matrices are recalculated exactly, the effect of the change of the CI vector on the orbitals is taken into account more accurately than in the Newton–Raphson method. This considerably improves convergence, particularly in the region far from the solution.

In the (step-restricted) AH method as proposed by Lengsfield[39,44,47] and

TABLE II
Convergence of CASSCF calculations for formaldehyde[a].

	Energy difference		
Iter.	Without coupling[b]	With coupling[c]	With coupling[d]
1	−0.095301120	−0.106133147	−0.105835733
2	−0.009714002	−0.001264146	−0.001561529
3	−0.001696277	−0.000000014	−0.000000044
4	−0.000446948		
5	−0.000141131		
6	−0.000052973		
7	−0.000022911		
8	−0.000010824		
9	−0.000005361		

[a] Active space: $3a_1$–$7a_1$, $1b_2$–$3b_2$, $1b_1$–$2b_1$, 3644 configurations; basis set and other details, see Ref. 55.

[b] In each iteration, Eq. (54) is solved with fixed CI coefficients. The CI coefficients are optimized with $\mathbf{R} = \mathbf{0}$ at the beginning of each iteration.

[c] $E^{(2)}(\mathbf{T}, \mathbf{c})$ fully optimized with respect to **T** and **c**.

[d] Same as footnote *c*, but only five updates of the CI coefficients in the first iteration.

used in various modifications by Shepard *et al.*[46], Golab *et al.*[52] and Jensen and Agren[70] the residual **Y** is calculated according to Eq. (60) for the special case $\mathbf{U} = \mathbf{1}$ only. (If the orbital CI coupling is included, further terms have to be added to **Y**.) After having obtained the solution $\mathbf{Y} = \mathbf{0}$ with $\mathbf{U} = \mathbf{1}$ in these methods, the next macro-iteration is started, i.e. a four-index transformation to obtain new operators $\mathbf{J}^{kl}$ and $\mathbf{K}^{kl}$ is necessary. In our method this transformation is only performed after several applications of the AH method to the energy approximation $E^{(2)}(\mathbf{T})$. This greatly reduces the number of four-index transformations and the overall effort. Tables II and III demonstrate the fast convergence for some CASSCF calculations. It is seen that convergence is reached in only 2–3 macro-iterations. Similar fast convergence behaviour has been observed in many other applications.

The total number of micro-iterations needed for solving the non-linear equations (53) is often fairly large. It is, therefore, important to make them as efficient as possible. In order to minimize the I/O time, the operators $\mathbf{J}^{kl}$ and $\mathbf{K}^{kl}$ should be kept in high-speed memory whenever possible, since their recovery from disc may be more expensive than their use in the calculation of **Y**. It is worth while to mention that we often found it advantageous to evaluate the operators $\mathbf{G}^{ij}$ as intermediate quantities. These operators only change if the CI coefficients are updated, and their calculation is particularly helpful if many micro-iterations are performed between CI updates. The matrix $\tilde{\mathbf{B}}$ is

TABLE III

Convergence of CASSCF calculations for various states of NO[a].

	Energy difference				
Iter.	$X^2\Pi$	$a^4\Pi$	$A^2\Sigma^+$	$B^2\Pi^b$	$b^4\Sigma^-$
1	−0.09099300	−0.06561940	−0.03737295	−0.06733878	−0.07296502
2	−0.00135266	−0.00930193	−0.00004712	−0.00282646	−0.00229034
3	−0.00000001	−0.00000786	−0.00000000	−0.00000012	−0.00000013

[a] Active space: 2σ–6σ, 1π–2π; 1σ, 2σ orbitals frozen. The starting orbitals were canonical SCF orbitals for the $^2\Sigma^+$ state in all cases. $R = 2.1$ bohr. Basis set:
Huzinaga[72] 11s, 7p (innermost 5s, 3p contracted), augmented on each atom by 2s, 1p, 2d functions with the following exponents:
N: s (0.051, 0.020); p (0.042); d (0.88, 0.22);
O: s (0.069, 0.027); p (0.053); d (1.2, 0.3);
Final energies are: −129.40697524, −129.12317041, −129.20572150, −129.08492834 and −129.14238817 hartree for the first to last columns, respectively.

[b] Second state of this symmetry optimized.

then obtained as

$$(\tilde{\mathbf{B}})_{ri} = (\tilde{\mathbf{A}})_{ri} + \sum_{s} U_{sr} \sum_{j} (\mathbf{G}^{ij}\mathbf{U}\Delta\mathbf{R})_{sj} \tag{73}$$

which requires about M^2N^2 operations. This is cheaper by a factor of $\frac{3}{2}M$ per micro-iteration than the application of Eq. (47). This saving often outweighs the additional cost for the calculation of the $\mathbf{G}^{ij}$ (Eq. (41), $\frac{3}{4}N^2M^4$ operations). The advantages are even greater if molecular symmetry can be employed, since only those blocks $(\mathbf{G}^{ij})_{rs}$ are needed in which (r, i) and (s, j) correspond to orbitals of the same symmetry. The evaluation of the operators $\mathbf{G}^{ij}$ is particularly efficient on vector computers, because it can be performed in terms of matrix multiplications with long vector lengths. In this case the elements G^{ij}_{rs}, P^{ij}_{kl} and Q^{ij}_{kl} (fixed ij) form vectors, and the operators J^{kl}_{rs} and K^{kl}_{rs} form supermatrices.

F. Optimization of Internal Orbital Rotations

The optimization method outlined in Sections II.C–E shows very fast and stable convergence behaviour when applied to CASSCF wavefunctions, in which all orbital rotations between occupied orbitals are redundant. However, experience has shown that convergence is often much slower when orbital rotations between strongly occupied valence orbitals have to be optimized. It has been shown that this is due to the fact that the energy approximation $E^{(2)}(\mathbf{T})$ is not invariant with respect to a unitary transformation between two doubly occupied orbitals[55]. If the 2 × 2 transformation in Eq. (34) is applied to a single Slater determinant with just two doubly occupied orbitals, the second-order energy $E^{(2)}(\mathbf{T})$ takes the form ($\alpha = R_{12}$, $T_{11} = T_{22} = \cos\alpha - 1$, $T_{12} = -T_{21} = \sin\alpha$)

$$E^{(2)}(\alpha) = e_1 + e_2(5 - 8\cos\alpha + 4\cos^2\alpha) \tag{74}$$

where

$$e_1 = 2(h_{11} + h_{22}) \tag{75}$$

$$e_2 = (11|11) + (22|22) + 4(11|22) - 2(12|12) \tag{76}$$

For small α, Eq. (74) can be approximated by

$$E^{(2)}(\alpha) = e_1 + e_2(1 + \tfrac{2}{3}\alpha^4 + \cdots) \tag{77}$$

Hence, $E^{(2)}(\mathbf{T})$ is not invariant with respect to α. In the Newton–Raphson approximation only terms up to second order in α are taken into account, and Eq. (77) shows that $E^{(2)}(\mathbf{R})$ has the correct invariance property. The implications of these findings are demonstrated in Fig. 4. This figure illustrates for a simple MCSCF wavefunction how the exact energy and the second-order energy $E^{(2)}(\mathbf{T})$ change as a function of the rotation angle α between two strongly occupied orbitals. Since one of these orbitals is correlated, the exact energy is not invariant with respect to α, but it is very flat. The energy approximation $E^{(2)}(\mathbf{T})$ shows the expected α^4 dependence. Therefore, the rotation angle predicted by minimization of $E^{(2)}(\mathbf{T})$ is much too small. In fact, too small step sizes for internal orbital rotations have been observed in many

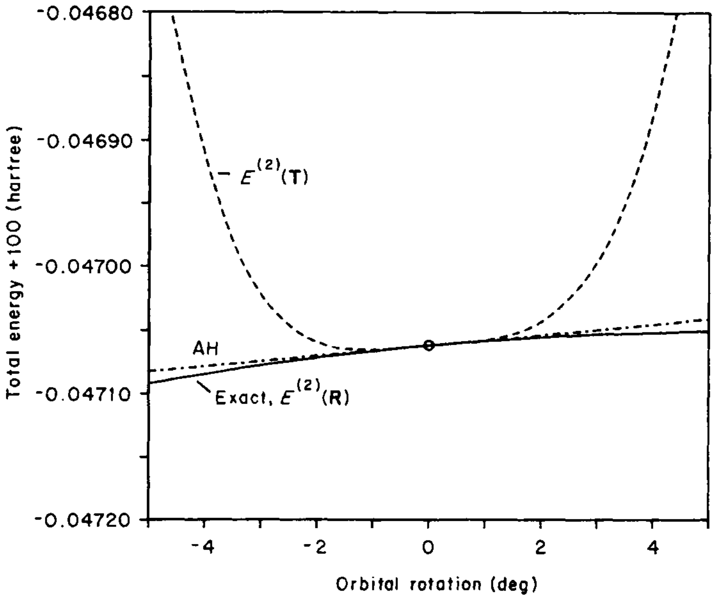

Fig. 4. The dependence of various energy approximations on the rotation 2σ–3σ for the HF molecule as in Fig. 1.

applications. The Newton–Raphson or undamped augmented Hessian methods predict steps which are of wrong direction or too large ($+61^\circ$ and -542°, respectively; the optimum angle is -53°). Therefore, the use of the NR or AH method often does not remedy convergence difficulties for such rotations.

The fact that the rotations between occupied orbitals cause most difficulties suggests a special treatment of these rotations. This is possible at rather modest expense, since for any unitary matrix $\mathbf{U}$ which transforms the internal orbitals only among themselves, the transformation of the internal one- and two-electron integrals h_{ij} and $(ij|kl)$ is very cheap and can be performed in high-speed memory. Hence, we can start each macro-iteration with an optimization of the internal–internal orbital rotations and the CI coefficients only. As for the minimization of $E^{(2)}(\mathbf{T},\mathbf{c})$ an uncoupled step-restricted augmented Hessian method can be used for this optimization. Since the number of internal–internal orbital rotations is small, the augmented orbital Hessian can be constructed and diagonalized explicitly. A CI update is done after 1–3 orbital optimization steps, which converge quadratically. Usually a total of 2–3 CI updates is sufficient. Since in each step the one- and two-electron integrals are transformed exactly, the process yields a variational energy which is fully optimized with respect to the parameter subspace considered.

Owing to the fact that the internal orbitals change in this optimization process, the operators h_{rs}, J^{kl}_{rs} and K^{kl}_{rs} also change. It is not necessary, however, to perform a second four-index transformation. Instead, the modified operators are obtained from the original ones by the much cheaper transformations

$$h_{rs} \rightarrow (\mathbf{U}^{+}\mathbf{h}\mathbf{U})_{rs} \tag{78}$$

$$J^{kl}_{rs} \rightarrow \sum_{ij} (\mathbf{U}^{+}\mathbf{J}^{ij}\mathbf{U})_{rs} U_{ik} U_{jl} \tag{79}$$

$$K^{kl}_{rs} \rightarrow \sum_{ij} (\mathbf{U}^{+}\mathbf{K}^{ij}\mathbf{U})_{rs} U_{ik} U_{jl} \tag{80}$$

The simplest method to perform this transformation requires about $\frac{3}{4}N^2M^4$ operations. Symmetry greatly reduces the effort. On vector machines the transformation can be performed in terms of matrix multiplications with long vector lengths (all $U_{ik}U_{jl}$ for a given k, l form a vector, the operators form supermatrices) and is therefore very fast. Using the transformed operators and orbitals, the optimization process is continued as described in Sections II.C–E.

The separate optimization of the internal orbital rotations at the beginning of each macro-iteration improves convergence considerably. However, this treatment so far neglects the coupling to the internal–external rotations. This coupling creates additional rotations R_{ij} between the internal orbitals when the non-linear equations (53) are solved. Convergence can be further improved

TABLE IV

Convergence of CASSCF calculations for the electronic ground state of CS_2[a].

	Energy difference		
Iter.	Without int. opt.[b]	With int. opt.[c]	With int. opt. and abs.[d]
1	−0.077490937	−0.085838896	−0.087484601
2	−0.005439995	−0.004644181	−0.003059391
3	−0.003146803	−0.000066851	−0.000006018
4	−0.002194213	−0.000000082	−0.000000000
5	−0.001326677		
6	−0.000661122		
7	−0.000240696		
8	−0.000047226		
9	−0.000002333		
10	−0.000000007		

[a] Geometry: $R = 3.1$ bohr, $\alpha = 105°$. Basis set:
C: Huzinaga[72] 8s, 4p, innermost 5s, 3p contracted; in addition one s (0.05), one p (0.05), and one d (0.5).
S: Huzinaga[72] 11s, 7p, innermost 6s, 4p contracted; in addition one s (0.05), one p (0.05), and one d (0.5).
Active space: $7a_1$–$10a_1$, $6b_2$–$8b_2$, $2b_1$–$3b_1$, $2a_2$ (3564 CSFs, 11 100 determinants). The $1a_1$–$5a_1$, $1b_2$–$4b_2$, $1b_1$ and $1a_2$ orbitals were canonical SCF orbitals and frozen; the $6a_1$ and $5b_2$ orbitals are doubly occupied and optimized.

[b] No extra optimization of internal–internal orbital rotations.

[c] Internal–internal orbital rotations fully optimized in the beginning of the second to fourth iteration.

[d] As in footnote *c*, but in addition one absorption step of elements R_{ij} in each iteration (see text).

TABLE V

Convergence of test calculations for the N_2 and CO molecules.

	Energy difference			
Iter.	N_2[a]	N_2[b]	CO[c]	CO[d]
1	−0.052436691	−0.053795845	−0.056168288	−0.114267390
2	−0.003387148	−0.002049568	−0.002605968	−0.006158512
3	−0.000023449	−0.000001875	−0.000001306	−0.000002589
4	−0.000000000	−0.000000000	−0.000000000	−0.000000000

[a] N_2 molecule; using IVOs as starting guess; basis set and configurations as in Ref. 51 and Table I. One absorption of elements R_{ij} per iteration.

[b] As in footnote *a*, but two absorptions of elements R_{ij} per iteration.

[c] CO molecule; using IVOs as starting guess; basis set and configurations as in Ref. 51, and Table I. One absorption of elements R_{ij} per iteration.

[d] As in footnote *c*, but canonical SCF orbitals used as starting guess.

by an iterative 'absorption' of these parameters R_{ij} into the orbital basis. This can be accomplished as follows. After solving the non-linear equations a unitary transformation matrix is formed from the elements $R_{ij} \simeq \frac{1}{2}(\mathbf{U} - \mathbf{U}^{\dagger})_{ij}$ (i, j internal only), and the transformations in Eqs (78)–(80) are repeated. The internal orbitals are transformed correspondingly. Then, some additional micro-iterations are necessary to solve the non-linear equations (53) and (58) with the new operators. In this case a good starting approximation is available by forming a unitary matrix from the previous R_{ai} and $R_{ij} = 0$. This process can be iterated until all R_{ij} remain zero when solving Eq. (53). In that case the internal orbital rotations have been treated to highest possible order with the operators $\mathbf{J}^{kl}$ and $\mathbf{K}^{kl}$ of the present macro-iteration. Since this is also true for the optimization of the CI coefficients, about the same convergence behaviour is expected for CASSCF (no internal–internal orbital rotations) and more general MCSCF calculations. In practice, we found it sufficient to absorb the internal–internal elements R_{ij} into the present orbital basis only once per macro-iteration. Furthermore, the initial optimization of the internal–internal orbital rotations is not necessary in the first macro-iteration.

Tables IV and V demonstrate the convergence improvements due to the internal optimization procedure. The first column of Table IV shows a calculation without the extra optimization of the internal orbital rotations. In the second column of Table IV the internal–internal rotations are optimized only at the beginning of the second and subsequent macro-iterations, but no absorptions of the elements R_{ij} have been performed. In the third column one absorption step has been done in each macro-iteration. The drastic reduction of the number of four-index integral transformations justifies the additional effort needed for the transformations in Eqs (78)–(80). The wavefunctions optimized in Table V are the same as those in Table I, and a comparison of these tables clearly demonstrates the convergence acceleration achieved with our method. The observed convergence behaviour is better than quadratic immediately from the first iteration even if bad starting orbitals are used. In the calculation for CO with SCF orbitals as a starting guess (last column in Table V), the initial orbital Hessian matrix had 21 negative diagonal elements, and probably a greater number of negative eigenvalues. In this case the initial 6σ and 2π orbitals were of Rydberg rather than of antibonding character. For N_2 it is shown in Table V that more than one absorption step in each iteration somewhat improves convergence. However, because it is usually not possible to save an iteration in this way, it is not recommended to perform more than one absorption step.

G. Treatment of Closed Shells

In many applications several orbitals are doubly occupied in all configurations of the MCSCF wavefunction. For such orbitals it is possible to simplify the algorithm, since the sub-blocks of the density matrices which involve closed-shell orbitals are of very simple structure. As shown below, they can be eliminated completely from the formalism.

The simplest way to deal with energetically low-lying closed-shell core orbitals is to take them directly from a preceding SCF calculation without further optimization. In this case one has to eliminate all rows and columns corresponding to core orbitals from the matrices $\mathbf{P}^{kl}, \mathbf{Q}^{kl}, \mathbf{R}, \mathbf{A}, \mathbf{B}$, etc., and replace the one-electron Hamiltonian $\mathbf{h}$ by a core Fock operator $\mathbf{F}^{c}$. This operator is calculated in the AO basis* according to

$$\tilde{\mathbf{F}}^{c} = \tilde{\mathbf{h}} + \tilde{\mathbf{G}}(\tilde{\mathbf{D}}^{c}) \tag{81}$$

where

$$(\tilde{\mathbf{D}}^{c})_{\mu\nu} = 2 \sum_{i(\text{closed})} X_{\mu i} X_{\nu i} \tag{82}$$

is the core first-order density matrix in the AO basis, and $\tilde{\mathbf{G}}(\tilde{\mathbf{D}})$ is defined as

$$\tilde{\mathbf{G}}(\tilde{\mathbf{D}})_{\mu\nu} = \sum_{\rho\sigma} \tilde{D}_{\rho\sigma}[(\mu\nu|\rho\sigma) - \tfrac{1}{2}(\mu\rho|\sigma\nu)] \tag{83}$$

The operator $\tilde{\mathbf{F}}^{c}$ has to be evaluated only once. It is transformed into the current MO basis at the beginning of each macro-iteration, i.e.

$$\mathbf{F}^{c} = \mathbf{X}^{\dagger} \tilde{\mathbf{F}}^{c} \mathbf{X} \tag{84}$$

Whenever the operator $\mathbf{G}(\mathbf{D})$ is used in the following, it will be assumed that it has been transformed into the MO basis similarly.

The freezing of core orbitals is usually a very good approximation if they are energetically well separated from the valence orbitals. However, the full optimization of all orbitals is sometimes desirable. This is the case, for instance, when the MCSCF calculation is followed by the evaluation of energy gradients with respect to the nuclear coordinates. It is, therefore, useful to consider explicitly the simplifications which are possible for closed-shell orbitals.

Using the anticommutation relations in Eqs (10)–(12) it is straightforward to derive the following expressions for the case that $|i\rangle$ is a closed-shell orbital:

$$D_{ij} = 2\delta_{ij}$$

$$P_{ij}^{kl} = 2\delta_{ij} D_{kl} - \tfrac{1}{2}(\delta_{il} D_{jk} + \delta_{ik} D_{jl}) \tag{85}$$

$$Q_{ij}^{kl} = 2\delta_{ik} D_{jl} - \tfrac{1}{2}(\delta_{il} D_{jk} + \delta_{ij} D_{kl}) \tag{86}$$

Using these relations, we obtain ($|i\rangle$ closed shell, $|j\rangle$ closed or open shell):

$$A_{ri} = 2(\mathbf{G}^{c})_{ri} \tag{87}$$

$$\mathbf{G}^{ij} = 2\delta_{ij} \mathbf{G}^{c} + \sum_{k} D_{jk} \mathbf{L}^{ik} \tag{88}$$

---

*In this section all quantities in the AO basis are marked with a tilde.

where

$$\mathbf{G}^{\mathrm{c}} = \mathbf{F}^{\mathrm{c}} + \sum_{kl(\mathrm{open})} D_{kl}(\mathbf{J}^{kl} - \tfrac{1}{2}\mathbf{K}^{kl}) \tag{89}$$

and

$$\mathbf{L}^{ik} = 4\mathbf{K}^{ik} - \mathbf{K}^{ki} - \mathbf{J}^{ik} \tag{90}$$

The summations in Eq. (89) run over open-shell orbitals only. For the case that $|i\rangle$ and $|j\rangle$ are open-shell orbitals, the same quantities are given by

$$A_{ri} = \sum_{j(\mathrm{open})} \left(F^{\mathrm{c}}_{rj} D_{ji} + \sum_{kl(\mathrm{open})} J^{kl}_{rj} P^{lk}_{ji} \right) \tag{91}$$

$$\mathbf{G}^{ij} = \mathbf{F}^{\mathrm{c}} D_{ij} + \sum_{kl(\mathrm{open})} (\mathbf{J}^{kl} P^{ij}_{kl} + 2\mathbf{K}^{kl} Q^{ij}_{kl}) \tag{92}$$

These formulae differ from Eqs (39) and (41) only in the use of $\mathbf{F}^{\mathrm{c}}$ instead of $\mathbf{h}$ and the restrictions of the summations. Hence, all second-order density matrix elements involving closed-shell orbitals have been eliminated. However, all operators $\mathbf{J}^{kl}$ and $\mathbf{K}^{kl}$ are still needed in Eq. (88). Since the computational effort for their evaluation depends strongly on the number of optimized orbitals, it would also be useful to eliminate the operators $\mathbf{J}^{kl}$ and $\mathbf{K}^{kl}$ involving any closed-shell orbitals. As shown in the following, this is possible in a direct MCSCF procedure.

In each micro-iteration, we have to evaluate

$$B_{ri} = A_{ri} + \sum_j (\mathbf{G}^{ij}\,\mathbf{T})_{rj} \tag{93}$$

or the similar quantity $\tilde{\mathbf{B}}$. For the case that $|i\rangle$ is a closed-shell orbital, we obtain

$$B_{ri} = 2(\mathbf{G}^{\mathrm{c}}\mathbf{U})_{ri} + \sum_j (\mathbf{L}^{ij}\mathbf{T}\mathbf{D})_{rj} \tag{94}$$

Defining the first-order change of the density matrix $\mathbf{D}$ as

$$\Delta\mathbf{D} = \mathbf{T}\mathbf{D} + \mathbf{D}\mathbf{T}^{\dagger} \tag{95}$$

Eq. (94) can be rewritten as

$$B_{ri} = 2(\mathbf{G}^{\mathrm{c}}\mathbf{U} + \mathbf{G}(\Delta\mathbf{D}))_{ri} \tag{96}$$

The operator $\mathbf{G}(\Delta\mathbf{D})$ describes the first-order change of $\mathbf{G}^{\mathrm{c}}$. It can be obtained directly from the two-electron integrals in the AO basis (cf. Eqs (83) and (84)). This requires transforming $\Delta\mathbf{D}$ into the AO basis:

$$\Delta\tilde{\mathbf{D}} = \mathbf{X}\Delta\mathbf{D}\mathbf{X}^{\dagger} \tag{97}$$

From Eq. (96) the relation of our method to a closed-shell Hartree–Fock procedure is apparent. For the case that there are only closed-shell orbitals, the variational conditions for $E^{(2)}(\mathbf{T})$ (Eq. (53)) take the form

$$(\mathbf{U}^{\dagger}\mathbf{B})_{ai} = 2(\mathbf{U}^{\dagger}\mathbf{F}^{\mathrm{c}}\mathbf{U} + \mathbf{U}^{\dagger}\mathbf{G}(\Delta\mathbf{D}))_{ai} = 0 \qquad \text{for all } a, i \tag{98}$$

In the usual first-order SCF procedure the second term accounting for the change of the Fock operator is neglected.

Next we have to consider the columns of **B** which correspond to open-shell orbitals ($|i\rangle$ open shell). These columns are given by

$$\begin{aligned} B_{ri} = {} & \sum_{j(\text{open})} (\mathbf{F}^{\text{c}}\mathbf{U})_{rj} D_{ji} + \sum_{jkl(\text{open})} [(\mathbf{J}^{kl}\mathbf{U})_{rj} P^{lk}_{ji} + 2(\mathbf{K}^{kl}\mathbf{T})_{rj} Q^{lk}_{ji}] \\ & + \sum_{j(\text{closed})} \sum_{k(\text{open})} D_{ik}(\mathbf{L}^{kj}\mathbf{T})_{rj} \end{aligned} \tag{99}$$

The first three terms on the right-hand side of Eq. (99) differ from Eq. (40) only by the restrictions on the summations and the replacement of **h** by $\mathbf{F}^{\text{c}}$. The last term accounts for the change of the Fock operator $\mathbf{F}^{\text{c}}$ caused by a variation of the closed-shell orbitals. It can be brought into the form

$$\sum_{j(\text{closed})} \sum_{k(\text{open})} D_{ik}(\mathbf{L}^{kj}\mathbf{T})_{rj} = (\mathbf{G}(\Delta\mathbf{D}^{\text{c}})\mathbf{D})_{ri} \tag{100}$$

where

$$\begin{aligned} (\Delta\mathbf{D}^{\text{c}})_{rs} &= (\mathbf{T}\mathbf{D}^{\text{c}} + \mathbf{D}^{\text{c}}\mathbf{T}^{\dagger})_{rs} \\ &= 2 \sum_{j(\text{closed})} (T_{rj}\delta_{js} + T_{sj}\delta_{rj}) \end{aligned} \tag{101}$$

is the first-order change of the closed-shell density matrix in the MO basis.

It follows from the above that the price one must pay for the elimination of the operators $\mathbf{J}^{kl}$ and $\mathbf{K}^{kl}$ involving closed-shell orbitals is the calculation of the two operators $\mathbf{G}(\Delta\mathbf{D})$ and $\mathbf{G}(\Delta\mathbf{D}^{\text{c}})$ from the two-electron integrals in the AO basis in each micro-iteration. If a large number of micro-iterations are necessary to solve the non-linear equations (53), this might be rather expensive. However, it is expected that energetically low-lying 'core orbitals' do not depend much on the changes of the valence orbitals and converge rapidly. It should, therefore, be possible to freeze all parameters R_{ri} which involve core orbitals in intermediate micro-iterations, and evaluate the operators $\mathbf{G}(\Delta\mathbf{D})$ and $\mathbf{G}(\Delta\mathbf{D}^{\text{c}})$ only a few times. Such an approximation is not appropriate, however, for closed-shell valence orbitals. For these orbitals the operators $\mathbf{L}^{ij}$ should be calculated explicitly.

In order to update the core orbitals the denominators D_{ri} in Eq. (63) are needed. Neglecting terms arising from the change of the operator $\mathbf{G}^{\text{c}}$, these can be approximated by

$$D_{ri} = 2(G^{\text{c}}_{rr} - G^{\text{c}}_{ii}) \tag{102}$$

Again, this approximation is not appropriate if $|i\rangle$ is a valence orbital.

H. A Direct Configuration-interaction Method for Complete Active Space Calculations

In the wavefunction optimization procedure outlined in the previous sections the coefficients $\{c_I\}$ can be optimized by any available CI procedure.

For short CI expansions one could simply construct and diagonalize the Hamiltonian matrices $\mathbf{H}^{(2)}$. For longer CI expansions a direct CI procedure must be employed. Here the desired eigenvectors are obtained iteratively without explicitly calculating and storing the Hamilton matrix. The basic operation in a direct CI iteration is the evaluation of the residual vectors*

$$\mathbf{y}^n = (\mathbf{H} - E^n)\mathbf{c}^n \tag{103}$$

for the required electronic states n. In terms of the molecular integrals and the coupling coefficients $\mathbf{y}$ can be written as

$$\begin{aligned} y_I = {} & \sum_{ij} h_{ij} \sum_J \langle \Phi_I | E_{ij} | \Phi_J \rangle c_J \\ & + \tfrac{1}{2} \sum_{ijkl} (ij|kl) \sum_J \langle \Phi_I | E_{ij,kl} | \Phi_J \rangle c_J - E c_I \end{aligned} \tag{104}$$

The coupling coefficients $\langle \Phi_I | E_{ij} | \Phi_J \rangle$ and $\langle \Phi_I | E_{ij,kl} | \Phi_J \rangle$ in principle need only be constructed once if they are stored on a formula tape. In our program this step can be performed by a graphical unitary group approach (GUGA)[27]. However, for large CASSCF calculations, the formula tape becomes exceedingly long. This limits the length of the configuration expansion which can be handled by this method to about 10^4 configurations. For longer CASSCF expansions the only remedy is to recalculate the coupling coefficients each time they are needed. Clearly, this requires a particularly efficient algorithm.

Recently, Siegbahn[73] proposed the use of the factorization[74–76]

$$\langle \Phi_I | E_{ij,kl} | \Phi_J \rangle = \sum_K \langle \Phi_I | E_{ij} | \Phi_K \rangle \langle \Phi_K | E_{kl} | \Phi_J \rangle - \delta_{jk} \langle \Phi_I | E_{il} | \Phi_J \rangle \tag{105}$$

where the summation runs over the full spin-eigenfunction basis. This factorization follows from the anticommutation relations in Eqs (10)–(12). The formula tape then need only contain the one-particle matrix elements $\langle \Phi_I | E_{ij} | \Phi_K \rangle$, ordered after the intermediate state label K. Even though this greatly reduces the number of stored coupling coefficients, the length of the formula tape can still be rather large. Hence, sorting and processing of the coupling coefficients may take considerable time.

A further development was made by Knowles and Handy[62]. They proposed to use Slater determinants instead of spin eigenfunctions as a basis $\{\Phi_I\}$. In this case the coupling coefficients $\langle \Phi_I | E_{ij} | \Phi_J \rangle$ take only the values ± 1 or 0, and can rapidly be recalculated each time they are required. With a suitable 'canonical' addressing scheme for the determinants and CI coefficients, the construction and use of the coupling coefficients can be vectorized. This makes it possible to use modern vector processors very efficiently.

The calculation of the residual vector is done in the following steps:

$$A^K_{kl} = \sum_J \gamma^{KJ}_{kl} c_J \tag{106}$$

---

*Here and in the following we omit the superscripts indicating a second-order approximation.