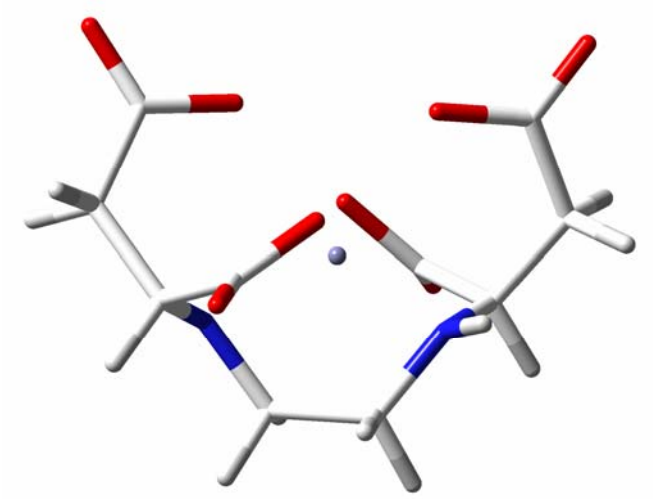
PhD Thesis

Optimization of Densities in Hartree-Fock and Density-functional Theory

Atomic Orbital Based Response Theory

and

Benchmarking for Radicals



Lea Thøgersen

Department of Chemistry

University of Aarhus

2005

"Experiments are the only means of knowledge at our disposal. The rest is poetry, imagination."

Max Planck

Contents

Preface

The present PhD thesis is the outcome of four years of PhD studies at the Faculty of Science, University of Aarhus, Denmark.

The thesis is divided into three distinct parts which can be read independently. Part 1 deals with the optimization of the one-electron density in Hartree Fock and density functional theory, and Part 2 deals with atomic orbital based response theory for Hartree Fock and density functional theory. Part 2 thus naturally follows after Part 1. In Part 3 benchmark results from FCI calculations on the radicals CN and CCH are given.

The work presented in Part 1 has resulted in papers I - III as listed in the following List of Publications and the work presented in Part 3 has resulted in papers V – VI. The work presented in Part 2 was initialized in the fall 2004 and will result in paper IV. The development of improved optimization algorithms for self-consistent field calculations is the subject on which I have spent the most of my time, and Part 1 therefore makes up the larger part of this thesis.

The work has been carried out under the supervision of and in collaboration with Dr. Jeppe Olsen and Professor Poul Jørgensen at the University of Aarhus. Some work was carried out during visits at The Royal Institute of Technology in Stockholm, Sweden, the University of Trieste, Italy and the University of Oslo, Norway. The following people have also contributed to the work presented in this thesis (see List of Publications): Paweł Sałek (The Royal Institute of Technology in Stockholm), Sonia Coriani (University of Trieste), Trygve Helgaker (University of Oslo), Stinne Høst (University of Aarhus), Danny Yeager (Texas A&M University), Andreas Köhn (University of Aarhus), Jürgen Gauss (University of Mainz), Péter Szalay (Eötvös Loránd University) and Mihály Kállay (University of Mainz).

The outline of the thesis is as follows: Part 1 is based on the published papers I – II and the unpublished paper III, but can be read independently of the papers. Certain discussions in the papers I - II are left out of the thesis and only referred to, as they might as well be read in the papers. Other discussions not published in the papers are presented in this thesis, including the latest developments of the algorithms. Part 2 is simply paper IV in preparation. Part 3 is based on the published papers V – VI and is basically a short version of paper V combined with selected results from paper VI. Also this part can be read independently of the papers.

List of Publications

This thesis includes the following papers. Number I, II, V and VI have already been published and are attached this thesis, whereas III and IV are in preparation.

Part 1

I. *The Trust-region Self-consistent Field Method: Towards a Black Box optimization in Hartree-Fock and Kohn-Sham Theories*,
L. Thøgersen, J. Olsen, D. Yeager, P. Jørgensen, P. Sałek, and T. Helgaker,
J. Chem. Phys. **121**, 16 (2004)

II. *The Trust-region Self-consistent Field Method in Kohn-Sham Density-functional Theory*,
L. Thøgersen, J. Olsen, A. Köhn, P. Jørgensen, P. Sałek, and T. Helgaker,
J. Chem. Phys. **123**, 074103 (2005)

III. *Augmented Roothaan-Hall for converging Densities in Hartree-Fock and Density-functional Theory*,
S. Høst, L. Thøgersen, P. Jørgensen and J. Olsen

Part 2

IV. *Atomic Orbital Based Response Theory*,
L. Thøgersen, P. Jørgensen, J. Olsen and S. Coriani

Part 3

V. *A Coupled Cluster and Full Configuration Interaction Study of CN and CN^-*,
L. Thøgersen and J. Olsen,
Chem. Phys. Lett. **393**, 36 (2004)

VI. *Equilibrium Geometry of the Ethynyl (CCH) Radical*,
P. G. Szalay, L. Thøgersen, J. Olsen, M. Kállay and J. Gauss,
J. Phys. Chem. A **108**, 3030 (2004).

Part 1

Improving Self-consistent Field Convergence

1.1 Introduction

The Hartree-Fock (HF) self-consistent field (SCF) method has been around in an orbital formulation since 1951, where it was introduced by Roothaan[1] and Hall[2], but today it is as significant as ever. Even though numerous higher correlated methods with superior accuracy have been developed since then, most of them still use the Hartree-Fock wave function as the reference function, and are thus still dependent on a functioning Hartree-Fock optimization. When Kohn and Sham[3] recognized in 1965 that the Roothaan-Hall SCF scheme had a lot to offer the density optimization in density functional theory (DFT), the DFT methods entered the chemical scene. Now it was in theory also possible to obtain results at the exact level from SCF calculations; if only the correct functional could be found. The developments in computer hardware and linear scaling SCF algorithms over the last decade have made it possible to carry out *ab initio* quantum chemical calculations on bio-molecules with hundreds of amino acids and on large molecules relevant for nano-science. Quantum chemical calculations are thus evolving to become a widespread tool for use in several scientific branches. It is therefore important that the algorithms work as black-boxes, such that the user outside quantum chemistry does not have to be concerned with the details of the calculations. Since no scientific results neither from the higher correlated calculations nor from the large-scale calculations can be achieved if the SCF optimization does not converge, it is necessary to take an interest in developing a sound, stable optimization scheme that can handle the complexity in the problems of the future.

This part of my thesis is a contribution to the quest for a black-box SCF optimization algorithm with optimal convergence properties. In Section 1.2, the basic Hartree-Fock/Kohn-Sham theory and notation of this part of the thesis is stated, and in Section 1.3 the efforts through the years to

improve the Roothaan-Hall SCF scheme are reviewed. Our contributions to the development of stable and physical sound SCF optimization schemes are presented in Section 1.4, and in Section 1.5 we study the quality of the schemes when applied for HF and DFT. Optimization of problems with several stationary points is discussed in Section 1.6, in Section 1.7 the scaling of the algorithms is accounted for, and Section 1.8 contains some convergence examples for HF and DFT calculations using the algorithms presented in Section 1.4. Finally, Section 1.9 contains concluding remarks; reviewing the results of this part of the thesis.

1.2 The Self-consistent Field Method

In the following we consider a closed-shell system with $N/2$ electron pairs. The basic theory of the Hartree-Fock (HF) and the Kohn-Sham (KS) density optimizations will be described simultaneously, and the differences will be noted as they appear. Since we are interested in extending the algorithms presented to large scale calculations, a formulation without reference to the delocalized molecular orbitals (MOs) is essential, and thus the focus will be on the density in the atomic orbital (AO) basis rather than the MOs themselves. All through the thesis, SCF will be used as a general term for HF and KS-DFT methods since they have the SCF optimization scheme in common. The orbital index convention used in this thesis is i, j, k, l for occupied MOs, a, b, c, d for virtual MOs, p, q for MOs in general, and Greek letters μ, ν, ρ, σ for AOs.

For closed-shell restricted Hartree-Fock or DFT, the electronic energy is given by

$$E_{\mathrm{SCF}} = 2\,\mathrm{Tr}\,\mathbf{hD} + \mathrm{Tr}\,\mathbf{DG}(\mathbf{D}) + h_{\mathrm{nuc}} + E_{\mathrm{XC}}(\mathbf{D}), \tag{1.1}$$

where $\mathbf{h}$ is the one-electron Hamiltonian matrix in the AO basis, h_{nuc} is the nuclear-nuclear repulsion contribution, and $\mathbf{D}$ is the (scaled) one-electron density matrix in the AO basis, $\mathbf{D} = \frac{1}{2}\mathbf{D}^{\mathrm{AO}}$, which satisfies the symmetry, trace, and idempotency conditions,

$$\begin{aligned} \mathbf{D}^{\mathrm{T}} &= \mathbf{D} \\ \mathrm{Tr}\,\mathbf{DS} &= \tfrac{N}{2} \\ \mathbf{DSD} &= \mathbf{D}, \end{aligned} \tag{1.2}$$

of a valid one-electron density matrix. $\mathbf{S}$ is the AO overlap matrix. The elements of $\mathbf{G}(\mathbf{D})$ are given by

$$G_{\mu\nu}(\mathbf{D}) = 2\sum_{\rho\sigma} g_{\mu\nu\rho\sigma} D_{\rho\sigma} - \gamma \sum_{\rho\sigma} g_{\mu\sigma\rho\nu} D_{\rho\sigma}, \tag{1.3}$$

where $g_{\mu\nu\rho\sigma}$ are the two-electron AO integrals. The first term in Eq. (1.3) represents the Coulomb contribution, and the second term is the contribution from exact exchange, with $\gamma = 1$ in Hartree-Fock theory, $\gamma = 0$ in pure DFT, and $\gamma \neq 0$ in hybrid DFT. The exchange-correlation energy $E_{\mathrm{XC}}(\mathbf{D})$ in Eq. (1.1) is a nonlinear and non-quadratic functional of the electronic density. This term is only

present in the energy expression for the DFT level of theory - the Hartree-Fock energy is expressed only by the first three terms of Eq. (1.1). The form of E_{XC} depends on the DFT functional chosen for the calculation.

The first derivative of the electronic energy with respect to the density is found as

$$\mathbf{E}^{(1)}_{SCF}(\mathbf{D}) = \frac{\partial E_{SCF}(\mathbf{D})}{\partial \mathbf{D}} = 2\mathbf{F}(\mathbf{D}), \tag{1.4}$$

where

$$\mathbf{F}(\mathbf{D}) = \mathbf{h} + \mathbf{G}(\mathbf{D}) + \tfrac{1}{2}\mathbf{E}^{(1)}_{XC}(\mathbf{D}) \tag{1.5}$$

is the Kohn-Sham matrix in DFT and, if the last term is excluded, the Fock matrix in Hartree-Fock theory. From now on $\mathbf{F}(\mathbf{D})$ is simply referred to as the Fock matrix. $\mathbf{E}^{(1)}_{XC}(\mathbf{D})$ is the first derivative of the term E_{XC} expanded in the density.

The Fock matrix is by design an effective one-electron Hamiltonian which is itself dependent on the eigenfunctions. Optimizing the electronic energy is thus a nonlinear problem and an iterative scheme must be applied. In 1951 Roothaan and Hall suggested an iterative procedure[1,2] in which a set of molecular orbitals (MOs) are constructed in each step through a diagonalization of the current Fock matrix, which in the AO formulation is written as

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon}, \tag{1.6}$$

where $\mathbf{S}$ is the AO overlap matrix, $\boldsymbol{\varepsilon}$ is a diagonal matrix containing the orbital energies, and the eigenvectors $\mathbf{C}$ contain the MO coefficients. The MOs, φ_p, are linear combinations of a finite set of one-electron basis functions, χ_μ, with $C_{\mu p}$ as expansion coefficients

$$\varphi_p = \sum_\mu \chi_\mu C_{\mu p}\ . \tag{1.7}$$

For the closed shell case the MOs can be divided into an occupied (φ^{occ}) and a virtual (φ^{virt}) part, where the occupied MOs each contain two electrons and the virtual orbitals are empty. If the *aufbau* ordering rule is applied, the occupied MOs are chosen as those with the lowest eigenvalues.

A new trial density $\mathbf{D}$ can then be constructed from the occupied orbitals as

$$\mathbf{D} = \mathbf{C}_{occ}\mathbf{C}^{T}_{occ}\ . \tag{1.8}$$

From this density a new Fock matrix can be evaluated from Eq. (1.5) and diagonalizing it according to Eq. (1.6) establishes the iterative procedure. The iterative cycle stops when self-consistency is obtained, that is, when the new density, energy or molecular orbitals do not change within some convergence threshold compared to the previous ones.

In an iterative scheme it is necessary to have a start guess. For the SCF case it should be a one electron density which fulfils Eq. (1.2), created directly or from a start guess of the molecular orbitals as in Eq. (1.8). Different approaches are used; a simple and easily applicable possibility is to obtain the starting orbitals by diagonalization of the one-electron Hamiltonian (H1-core). This is the start guess most widely used in this thesis since it is always available. Another popular possibility is to create a semi-empirical start guess where the orbitals resulting from a semi-empirical calculation (e.g. Hückel) on the molecule are fitted to the current basis.

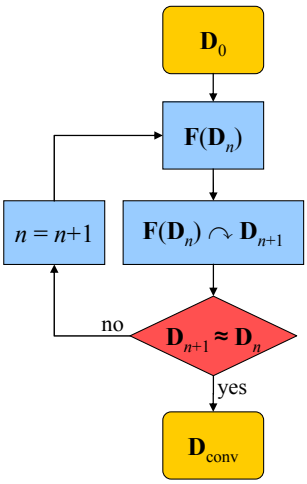


Fig. 1.1 Flow diagram of the SCF scheme.

The steps of the self-consistent field (SCF) scheme are summarized from the density point of view in Fig. 1.1: From a density matrix start guess a Fock matrix is constructed. From this Fock matrix a new density matrix can be found and so an iteration procedure is established which continues until self consistency. The step creating a new density from a Fock matrix will be referred to as the Roothaan-Hall (RH) step throughout this thesis, regardless if it is a diagonalization of the Fock matrix or some alternative scheme.

The purpose of an SCF optimization is typically to find the global minimum. Since the HF/KS equations are nonlinear, several stationary points might exist, and depending on the start guess and the optimization procedure, the converged result can be representing a local minimum as well as a global or even a saddle point. By evaluating the lowest Hessian eigenvalue it can be realized whether the stationary point is a minimum or a saddle point, but no simple test can reveal whether a minimum is global or not. The use of the term “convergence” in this thesis will simply refer to the iterative development from the start guess to a self-consistent density with a gradient below the convergence threshold. The issues connected with problems where several stationary points can be found are discussed in Section 1.6.

Since Roothaan and Hall suggested the iterative diagonalization procedure as a means to solve the Hartree-Fock equations and Kohn and Sham suggested using the same scheme for optimizing the electron density for density functional theory[3], the SCF methods have been used extensively in quantum chemistry. Unfortunately, it turned out that the simple fixed point scheme sketched in Fig. 1.1 converges only in simple cases. Already around 1960 it was recognized that the method sometimes fails to converge and that divergent behavior in some cases is intrinsic[4,5].

1.3 A Survey of Methods for Improving SCF Convergence

Numerous suggestions have been made to improve upon the convergence of Roothaan and Hall's original scheme or to replace it with an alternative scheme. The suggestions can be crudely divided into three different categories; energy minimization, damping/extrapolation, and level shifting. Furthermore the different suggestions in these categories have been combined in various ways. The two latter categories are modifications to the Roothaan-Hall scheme, whereas energy minimization is a means of avoiding the iterative diagonalization scheme and instead use some optimization scheme on an energy function.

To my knowledge these categories embrace all convergence improvements suggested over the years, except for the method of fractionally occupying orbitals around the Fermi level[6] which does not fit in any of the categories. As mentioned, the start guess has a great impact on the optimization, and a poor start guess with the wrong electron configuration can use many iterations changing to a more optimal electron configuration and in some cases the proper electron configuration is never found and the calculation diverges. In the methods using fractional occupations, a number of orbitals around the Fermi level are allowed to have non-integral occupation. The non-integral occupations are determined from the Fermi-Dirac distribution which is a function of the temperature. The non-integral occupations are updated in each iteration, and corrected such that the total number of electrons is constant. During the optimization either the temperature is decreased to T = 0K or the number of orbitals allowed to have non-integral occupation is decreased, to have only integer occupations at the end of the optimization. It is thus possible to optimize the electron configuration in an effective manner in the beginning of the SCF optimization, and when the proper configuration has been found, the rest of the optimization has a better chance of convergence since the start guess in a way has been improved.

In the following, the focus will be on the efforts to improve the convergence behavior of the SCF scheme through optimization algorithm development in the three categories listed above. Other efforts bear as much significance and should also be acknowledged, in particular should be mentioned the generalizations of many well-functioning schemes to the unrestricted level of theory which has its own challenges. Also the quest for construction of an improved start guess is important. It is obvious that with an improved start guess, less is demanded from the optimization method and thus some convergence problems inherent in the methods could be avoided. In the last decade the effort in SCF scheme development has for a large part been put in decreasing the scaling of the methods to allow calculations on larger molecules. Scaling is a very important subject and it should not be ignored. Section 1.7 will therefore discuss the scaling of the algorithms presented in

this thesis. Despite the importance of these three SCF related subjects, the rest of this section will be almost solely on efforts to improve convergence through optimization algorithm development.

1.3.1 Energy Minimization

One of the problems in the simple Roothaan-Hall procedure is the lack of guarantees for energy decrease in the iterative steps. This was pointed out by McWeeny, and he thus introduced a steepest descent procedure[7,8] as an energy minimization alternative to Roothaan and Hall's repeated diagonalizations. Steepest descent optimizations have the benefit that a decrease in energy can be guaranteed for each step. McWeeny's scheme suffers, however, from a slow convergence rate[5] as often seen for steepest descent methods. Fletcher and Reeves proposed the conjugate gradient optimization method[9] instead, which often is more efficient than steepest descent and is guaranteed to converge in a number of steps equal to the dimension of the problem.

A decade later Hilliers and Saunders suggested an improvement to the McWeeny scheme called energy-weighted steepest descent[10], in which the coordinates in the orbital space are energy-weighted. In 1976 this work was generalized by Seeger and Pople. They realized that another problem in the simple Roothaan procedure is the possibility for discontinuous changes in the orbitals which do not necessarily lower the energy. To ensure energy descent it is necessary to be able to follow such changes continuously, and methods like the steepest descent have the possibility to do so. Their procedure proceeds in small steps, where the new occupied trial orbitals are selected based on a criterion of overlap with the previous set. This technique ensures stability and avoids switching of orbital occupation. The step is found by a univariate search[11] in the energy, on a path that passes through the point corresponding to the next iteration step of the classical procedure. Their scheme can therefore also be seen as a polynomial interpolation along a path joining successive SCF cycles. Half a decade later, Camp and King followed the same strategy of a univariant cubic fit technique[12], but with a different parameterization. Stanton also suggested a similar approach[13], but whereas the Seeger-Pople approach requires the evaluation of the Fock matrix at interior points on the interpolative path, Stanton's scheme uses a cubic interpolation, where only the end point properties are needed, making it a less expensive method.

Another way of improving the convergence properties is to evaluate the gradient and Hessian of the electronic energy analytically with respect to some variational parameter, and then optimize the energy through Newton-Raphson steps resulting in a quadratically convergent[14] scheme, at least in the region close to the optimized state where a second order approximation is reasonable. These methods are computationally very expensive since a four index transformation is required to obtain the Hessian information. In 1981 Bacskay proposed a quadratically convergent SCF (QC-SCF) method[15] which escapes the four index transformation while requiring four or five micro iterations

per step (in non-problematic cases), each of which is about as expensive computationally as building a Fock matrix. His method was inspired from single excitation configuration interaction (SX-CI) and multi-configurational SCF (MC-SCF). A possible divergence of the scheme can be overcome by moderating the orbital update step by the augmented Hessian method[16] or trust radius techniques[17]. Even though it is still quite expensive, the method is also used today for cases with convergence problems, since a decrease in energy can be ensured step by step and it has quadratic convergence properties near the optimized state.

Around 1995, the interest for linear scaling SCF methods took on, since the development in computer hardware had made calculations on large molecules possible. With newly developed algorithms the evaluation of the Fock matrix, with the formal scaling of N^4 arising from the four-index integrals, could now routinely be decreased to a near-linear scaling. The diagonalization with a N^3 scaling in standard Roothaan-Hall was now the bottle neck. Inspiration was found in tight binding theory[18-20], where a number of linear scaling approaches had been suggested earlier[21]. To obtain linear scaling of the RH step it is necessary to avoid the diagonalization and to ensure sparsity in the matrices. This is a problem since the convenient canonical MO basis is inherently delocalized. Some of the well known schemes were reformulated in localized MOs[22], while others developed strict AO formulations[20,23-25]. Most of the suggested linear scaling methods did not arise so much to improve convergence as to improve the scaling, and will therefore not be discussed in further detail.

Very recently Francisco, Martínez and Martínez introduced their globally convergent trust region methods for SCF[26], where the standard fixed-point Roothaan-Hall step is replaced by a trust region optimization of a model energy function. This algorithm has very nice features since it can be proved to be globally convergent, and the step sizes are controlled dynamically through a trust region update scheme. The convergence rate seems rather random though; sometimes perfect and sometimes hopeless, but only small test examples have been published, so time will show.

1.3.2 Damping and Extrapolation

In his SCF study of atoms, Hartree noted convergence difficulties and suggested a so-called damping scheme[27] as a modification to the iterative procedure. Instead of using the newly constructed density $\mathbf{D}_{n+1}$, which corresponds to a full step, a linear combination of the new density matrix with the previous one is constructed

$$\mathbf{D}_{n+1}^{\text{damp}} = \mathbf{D}_n + \lambda(\mathbf{D}_{n+1} - \mathbf{D}_n) = \lambda\mathbf{D}_{n+1} + (1-\lambda)\mathbf{D}_n \ , \tag{1.9}$$

where λ – the damping factor - is a scalar chosen between zero and one. The iterative sequence is then continued with $\mathbf{D}^{\text{damp}}$ as the new density. Hartree found that this scheme could force convergence in problematic cases.

To get an idea of the effect of the damping factor, we consider a block-diagonal Fock matrix in the MO basis

$$\mathbf{F}^{\text{MO}} = \begin{pmatrix} \boldsymbol{\varepsilon}_{\text{o}} & \mathbf{F}_{\text{ov}} \\ \mathbf{F}_{\text{vo}} & \boldsymbol{\varepsilon}_{\text{v}} \end{pmatrix}, \tag{1.10}$$

where 'o' denotes occupied, 'v' virtual and $[\boldsymbol{\varepsilon}_{\text{o}}]_{ij} = \delta_{ij}\varepsilon_i$ and $[\boldsymbol{\varepsilon}_{\text{v}}]_{ab} = \delta_{ab}\varepsilon_a$. The change in electronic energy from the first order variation of the occupied orbitals through first-order perturbation theory is then given as

$$\Delta E_{\text{SCF}}^{(1)} = 4 \sum_{a}^{\text{virtual}} \sum_{i}^{\text{occupied}} \frac{-F_{ai}^2}{(\varepsilon_a - \varepsilon_i)} \,. \tag{1.11}$$

If this first order term is negative and sufficiently small such that the higher order contributions are insignificant, then a decrease in the electronic energy is seen. If the MOs obey the *aufbau* principle, then all $\varepsilon_i < \varepsilon_a$ and it is clear that the term is negative as desired. The Hartree damping of Eq. (1.9) roughly corresponds to multiplying the numerator of Eq. (1.11) by the factor λ, which is positive and less than one

$$\Delta E_{\text{SCF}}^{(1)} = 4 \sum_{a}^{\text{virtual}} \sum_{i}^{\text{occupied}} \frac{-\lambda F_{ai}^2}{(\varepsilon_a - \varepsilon_i)} \,, \tag{1.12}$$

thus giving the opportunity to obtain a negative first order change of arbitrarily small magnitude, making the higher order terms insignificant. Though this would seem promising, the *aufbau* principle is seldom obeyed all through the optimization.

If λ could be freely chosen, the damping technique would lead to an extrapolation scheme in the densities. Since SCF generates an iterative sequence where each step only depends upon the preceding, it was natural to apply the mathematical extrapolation methods (e.g. the Aitken extrapolation[28] procedures) on SCF to improve in particular the convergence rate close to the minimum. When the individual MO expansion coefficients are chosen as the extrapolated parameters, as Winter and Dunning Jr.[29] suggested, unphysical result may be obtained, though they can be corrected at the end of the calculation. Nielsen used instead the density matrix as the extrapolated parameter[30] and an eigenvalue extrapolation instead of the Aitken method. This led to a scheme more similar to Hartree damping, but with λ found within the eigenvalue extrapolation scheme.

Different approaches have been taken to dynamically find the damping factor λ. Zerner and Hehenberger[31] found it based on an extrapolation of the Mulliken gross population. Karlström[32] expressed the electronic energy in the damped density $E(\mathbf{D}^{\text{damp}})$ and used the first derivative with respect to λ, to choose in each iteration the λ that minimized the electronic energy.

None of these schemes were very successful solving the convergence problems. They all had some particular problematic cases they could handle better than the predecessors, but in general they did not catch on. Pulay then suggested in the early 1980s to use the norm of a linear combination of error vectors $\mathbf{e}_i$ from the individual iterations, where the vanishing of the error vector is a necessary and sufficient condition for SCF convergence. The norm is then optimized with respect to the coefficients c_i

$$\tilde{e}(\mathbf{c}) = \left| \sum_{i=1}^{n} c_i \mathbf{e}_i \right|, \tag{1.13}$$

where n is the number of previous iterations, and the coefficients are restricted to add up to 1

$$\sum_{i=1}^{n} c_i = 1. \tag{1.14}$$

The resulting coefficients are used to construct a favorable linear combination of the previous Fock matrices

$$\bar{\mathbf{F}} = \sum_{i=1}^{n} c_i \mathbf{F}_i , \tag{1.15}$$

which is diagonalized to obtain a new density, and so the iterative procedure is reestablished. This was the first density subspace minimization scheme that deliberately exploited the information obtained in the previous iterations and he named the approach DIIS[33] for "Direct Inversion in the Iterative Subspace". For the special case of two matrices, the DIIS density corresponds to the damped density of Eq. (1.9), but with no restrictions on λ. A decade later the DIIS algorithm was a standard option in most *ab initio* programs and had effectively solved a number of the convergence problems. The orbital rotation gradient was typically used as the error vector for wave function optimizations, and Sellers pointed out[34] that the DIIS algorithm exploits the second-order information contained in a set of gradients to obtain quadratic convergence behavior. Some numerical problems were seen though, where numerical instabilities appeared because of linear dependencies in the space of error vectors. Sellers introduced the C2-DIIS method[34], which is similar to DIIS except the restriction is on the squares of the coefficients

$$\sum_{i=1}^{n} c_i^2 = 1, \tag{1.16}$$

with a renormalization at the end. This gives an eigenvalue problem to be solved instead of the set of linear equations in normal DIIS, and thus singularities are more easily handled. However, one of the examples (Pd_2 in the Hyla-Kripsin basis set[35]) given in ref. [34], where DIIS supposedly diverges, converges for our plain DIIS implementation to 10^{-7} in the energy in 14 iterations.

Even though DIIS is successful, examples of divergence with no relation to numerical instabilities have been encountered over the years. In the year 2000 Cancès and Le Bris presented a damping algorithm named the Optimal damping Algorithm[36] (ODA) that ensures a decrease in energy at each iteration and converges toward a solution to the HF equations. In ODA the damping factor λ is found based on the minimum of the Hartree-Fock energy for the damped density in Eq. (1.9)

$$\begin{aligned} E_{\mathrm{HF}}\left(\mathbf{D}_{n+1}^{\mathrm{damp}},\lambda\right) = E_{\mathrm{HF}}\left(\mathbf{D}_n\right) + 2\lambda\,\mathrm{Tr}\,\mathbf{F}\left(\mathbf{D}_n\right)\left(\mathbf{D}_{n+1}-\mathbf{D}_n\right) \\ +\lambda^2\,\mathrm{Tr}\left(\mathbf{D}_{n+1}-\mathbf{D}_n\right)\mathbf{G}\left(\mathbf{D}_{n+1}-\mathbf{D}_n\right) + h_{\mathrm{nuc}} \quad , \end{aligned} \tag{1.17}$$

much like Karlström did it in 1979. The damping factor is thus optimized in each iteration, hence the name of the algorithm.

Recently Kudin, Scuseria, and Cancès proposed a method in which the gradient-norm minimization in DIIS is replace by a minimization of an approximation to the true energy function and they named it the energy DIIS (EDIIS) method[37]. Where the ODA used the energy expression of Eq. (1.17) to find the optimal λ, EDIIS uses an approximation of the Hartree-Fock energy for the averaged density

$$\bar{\mathbf{D}} = \sum_{i=1}^{n} c_i \mathbf{D}_i \ , \tag{1.18}$$

$$E^{\mathrm{EDIIS}}\left(\bar{\mathbf{D}},\mathbf{c}\right) = \sum_{i=1}^{n} c_i E_{\mathrm{SCF}}\left(\mathbf{D}_i\right) - \tfrac{1}{2}\sum_{i,j=1}^{n} c_i c_j\,\mathrm{Tr}\left(\left(\mathbf{F}_i-\mathbf{F}_j\right)\cdot\left(\mathbf{D}_i-\mathbf{D}_j\right)\right), \tag{1.19}$$

where the sum of the coefficients c_i is still restricted to 1. They combine the scheme with DIIS, such that the EDIIS optimized coefficients are used to construct the averaged Fock matrix if all coefficients fall between 0 and 1. If not, the coefficients from the DIIS scheme are used instead. The EDIIS scheme introduces some Hessian information not found in DIIS and thus improves convergence in cases where the start guess has a Hessian structure far from the optimized one. For non-problematic cases and near the optimized state EDIIS has a slower convergence rate than DIIS, but it has been demonstrated that EDIIS can converge cases where DIIS diverges.

Recently, we suggested another subspace minimization algorithm along the same line as EDIIS, but with a smaller idempotency error in the energy model and the same orbital rotation gradient in the subspace as the SCF energy (the EDIIS energy model actually has a different gradient). We named it TRDSM[38] for trust region density subspace minimization since a trust region optimization is

carried out of the energy model in the subspace of previous densities. In the second paper on TRDSM[39], a comparison with the EDIIS and DIIS models can be found stating explicitly that the EDIIS energy model does not have the correct gradient and is wrong for other reasons as well at the DFT level of theory.

Many of the energy minimization techniques can be combined with a damping or extrapolation scheme to improve the convergence. Typically, DIIS has been the choice[24,40,41], but TRDSM could be used just as well.

1.3.3 Level Shifting

In 1973 Saunders and Hillier introduced the level shift concept[42]. They suggested adding a positive scalar μ to the diagonal of the virtual-virtual block of the Fock matrix in the MO basis, Eq. (1.10), before diagonalizing

$$\left(\mathbf{F}^{\mathrm{MO}} + \mu\left(\mathbf{I} - \mathbf{D}^{\mathrm{MO}}\right)\right)\mathbf{C} = \mathbf{C}\boldsymbol{\varepsilon}\,, \tag{1.20}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{D}^{\mathrm{MO}}$ is the scaled one-electron density matrix in the MO basis with 1 in the diagonal of the occupied-occupied block and zeros for the rest.

To compare level shifting with the damping scheme of Hartree[27], consider the first order variation in the energy change as in Eq. (1.11); the level shift μ then corresponds to adding a positive constant to the denominator

$$\Delta E_{\mathrm{SCF}}^{(1)} = 4\sum_{a}^{\text{virtual}}\sum_{i}^{\text{occupied}} \frac{-F_{ai}^2}{\left(\varepsilon_a - \varepsilon_i + \mu\right)}\,. \tag{1.21}$$

The level shift thus has, as the damping factor, the possibility to decrease the magnitude of the term. The problems with respect to the *aufbau* principle mentioned in connection with the damping can be overcome with the level shift. The level shift can separate the occupied orbitals from the virtuals and thereby ensure a positive denominator and an overall decrease in energy. As the level shift is increased towards infinity, the obtained decrease in energy will correspond to that of the steepest descent method as explained in Section 1.4.1.4, and thus the convergence will be slow. This connection between a large gap between the occupied and the virtual orbitals (HOMO-LUMO gap) and slow convergence was exploited by Bhattacharyya in 1978 to accelerate convergence for cases with large HOMO-LUMO gaps. His “reverse level shift” technique[43] uses a negative level shift instead of a positive, thus decreasing the gap and accelerating the convergence.

In 1977, Carbó, Hernández and Sanz claimed unconditional convergence for an SCF process with a properly used level shift[44], and two decades later, Cancès and Le Bris[45] made a formal proof that for

any initial guess $\mathbf{D}_0$, there exists a level shift $\mu_0 > 0$ such that for level shift parameters $\mu > \mu_0$, the energy decreases at each step and converges towards a stationary value.

The level shift technique is still routinely used for cases where the DIIS scheme has problems. The level shifts are typically found on a trial and error basis. Recently, we advocated the use of a level shift to control the changes introduced in the Roothaan-Hall step[38], and we suggested a way of optimizing the level shift at each iteration based on physical arguments and without guesswork. The algorithm is based on the trust region philosophy in which a model energy function is optimized, but restricted with respect to the step length. We thus named the algorithm trust region Roothaan-Hall (TRRH), even though it is not a true trust region optimization scheme like e.g. the energy minimization of Francisco, Martínez, and Martínez[26] or our TRDSM scheme[38].

Level shifting can be combined with a damping or extrapolation scheme. When the TRRH approach is combined with the subspace minimization method TRDSM it seems to outperform DIIS in stability and to have a better or similar convergence rate, as will be illustrated in the following sections. Combining level shifting with DIIS can occasionally be a benefit, but typically DIIS and level-shifting does not work well together, and in Section 1.4.1.3 we will try to justify this.

1.4 Development of SCF Optimization Algorithms

The SCF scheme as it typically looks today is sketched in Fig. 1.2. Compared to Fig. 1.1, the step ② is inserted, illustrating a density subspace minimization, where some function f is minimized with respect to the coefficients c_i which expand the previous densities $\mathbf{D}_i$. The function f could be the gradient norm as in DIIS or some energy model approximating the SCF energy in the subspace of the previous densities as in EDIIS and TRDSM. In the Roothaan-Hall step ①, the averaged Fock matrix $\bar{\mathbf{F}}$ found from the optimization in ② is then used instead of the most recent Fock matrix $\mathbf{F}(\mathbf{D}_n)$ to find a new trial density $\mathbf{D}_{n+1}$. In general, the averaged density matrix $\bar{\mathbf{D}}$ is not idempotent and therefore does not represent a valid density matrix; moreover, since the Kohn-Sham matrix (unlike the Fock matrix) is nonlinear in the density matrix, the averaged Kohn-Sham matrix $\bar{\mathbf{F}}$ is different from $\mathbf{F}(\bar{\mathbf{D}})$. For these reasons, the averaged Fock matrix $\bar{\mathbf{F}}$ cannot be associated uniquely with a valid Fock matrix. Usually, this does not matter much since the subsequent diagonalization of the Fock matrix nevertheless produces a valid density matrix

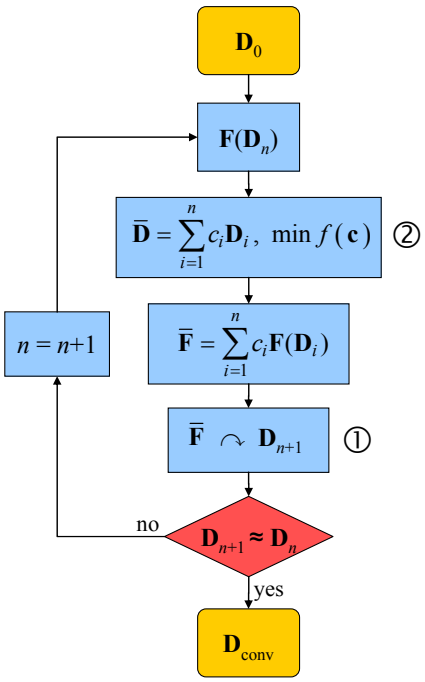


Fig. 1.2 Flow diagram of the SCF scheme including the density subspace minimization step.

according to Eq. (1.8). The complications arising from the use of the averaged Fock matrix is disregarded in the following, noting that the errors introduced by this approach may easily be corrected for, if necessary.

The rest of this part of the thesis will focus on the work we have done over the last couple of years to improve SCF convergence. We have made developments in all of the three categories of the previous section. The density subspace minimization scheme TRDSM and the level shift scheme in TRRH, both briefly described in the previous section, make up a total scheme we have named TRSCF, where each SCF iteration contains a TRDSM and a TRRH step. The first subsection will go into further detail on TRRH and will thus be concerned with our modifications to step ① in Fig. 1.2. The second subsection will likewise go into further detail on TRDSM and will describe the scheme we apply in step ②. In the third subsection, a recently developed energy minimization procedure will be presented. The procedure merges step ① and ② integrating a subspace minimization in the optimization of a new trial density.

This section will primarily take the Hartree-Fock point of view, acknowledging that with small adjustments and the word Fock replaced by Kohn-Sham, it would describe the DFT situation as well. In Section 1.5 the differences appearing when the algorithms are applied to the HF and DFT cases, respectively, will be discussed.

1.4.1 Dynamically Level Shifted Roothaan-Hall

The problems inherent to the RH diagonalization method are the discontinuous changes in the density and the lack of guarantees for energy decrease. To overcome these problems, we introduced in 2004 a means to restrict the RH step to the trust region of the RH energy model, with the purpose of both controlling the changes in the density and ensuring an energy decrease. Since then, the same ideas have been put forward by Francisco *et. al.*[26] as well, suggesting a trust region optimization of a RH energy model.

In this section, our trust region Roothaan-Hall scheme and related subjects are discussed. In particular, we present two different schemes for dynamic level shifting and an alternative to diagonalization.

1.4.1.1 RH Step with Control of Density Change

The solution of the traditional Roothaan–Hall eigenvalue problem Eq. (1.6) may be regarded as the minimization of the sum of the energies of the occupied MOs[8,46]

$$E^{\mathrm{RH}}(\mathbf{D}) = 2\sum_i \varepsilon_i = 2\,\mathrm{Tr}\,\mathbf{F}_0\mathbf{D} \tag{1.22}$$

subject to MO orthonormality constraints

$$\mathbf{C}_{\mathrm{occ}}^{\mathrm{T}}\mathbf{S}\mathbf{C}_{\mathrm{occ}} = \mathbf{I}_{N/2}\,, \tag{1.23}$$

where $\mathbf{F}_0$ is typically obtained as a weighted sum of the previous Fock matrices such as $\overline{\mathbf{F}}$ in Eq. (1.15). Since Eq. (1.22) represents a crude model of the true Hartree-Fock energy (with the same first-order term, but different zero- and second-order terms), it has a rather small trust radius. A global minimization of $E^{\mathrm{RH}}(\mathbf{D})$, as accomplished by the solution of the Roothaan–Hall eigenvalue problem Eq. (1.6), may therefore easily lead to steps that are longer than the trust radius and hence unreliable. To avoid such steps, we shall impose on the optimization of Eq. (1.22) the constraint that the new density matrix $\mathbf{D}$ does not differ much from the old $\mathbf{D}_0$, that is, the S-norm of the density difference should be equal to a small number Δ

$$\|\mathbf{D}-\mathbf{D}_0\|_{\mathrm{S}}^2 = \mathrm{Tr}\,(\mathbf{D}-\mathbf{D}_0)\mathbf{S}(\mathbf{D}-\mathbf{D}_0)\mathbf{S} = -2\,\mathrm{Tr}\,\mathbf{D}_0\mathbf{S}\mathbf{D}\mathbf{S} + N = \Delta\,, \tag{1.24}$$

where N is the number of electrons – see Eq. (1.2) – and the S-norm used throughout this thesis is defined as

$$\|\mathbf{A}\|_{\mathrm{S}}^2 = \mathrm{Tr}\,\mathbf{A}\mathbf{S}\mathbf{A}\mathbf{S} \tag{1.25}$$

for symmetric $\mathbf{A}$. The optimization of Eq. (1.22) subject to the constraints Eq. (1.23) and Eq. (1.24) may be carried out by introducing the Lagrangian

$$L = 2\,\mathrm{Tr}\,\mathbf{F}_0\mathbf{D} - 2\mu\left(\mathrm{Tr}\,\mathbf{D}\mathbf{S}\mathbf{D}_0\mathbf{S} - \tfrac{1}{2}(N-\Delta)\right) - 2\,\mathrm{Tr}\,\boldsymbol{\eta}\left(\mathbf{C}_{\mathrm{occ}}^{\mathrm{T}}\mathbf{S}\mathbf{C}_{\mathrm{occ}} - \mathbf{I}_{N/2}\right), \tag{1.26}$$

where μ is the undetermined multiplier associated with the constraint Eq. (1.24), whereas the symmetric matrix $\boldsymbol{\eta}$ contains the multipliers associated with the MO orthonormality constraints. Differentiating this Lagrangian with respect to the MO coefficients and setting the result equal to zero, we arrive at the level-shifted Roothaan–Hall equations:

$$(\mathbf{F}_0 - \mu\mathbf{S}\mathbf{D}_0\mathbf{S})\tilde{\mathbf{C}}_{\mathrm{occ}}(\mu) = \mathbf{S}\tilde{\mathbf{C}}_{\mathrm{occ}}(\mu)\boldsymbol{\lambda}(\mu)\,. \tag{1.27}$$

Since the density matrix, Eq. (1.8), is invariant to unitary transformations among the occupied MOs in $\tilde{\mathbf{C}}_{\mathrm{occ}}(\mu)$, we may transform this eigenvalue problem to the canonical basis:

$$(\mathbf{F}_0 - \mu\mathbf{S}\mathbf{D}_0\mathbf{S})\mathbf{C}_{\mathrm{occ}}(\mu) = \mathbf{S}\mathbf{C}_{\mathrm{occ}}(\mu)\boldsymbol{\varepsilon}(\mu), \tag{1.28}$$

where the diagonal matrix $\boldsymbol{\varepsilon}(\mu)$ contains the orbital energies. Note that, since $\mathbf{D}_0\mathbf{S}$ projects onto the part of $\mathbf{C}_{\mathrm{occ}}$ that is occupied in $\mathbf{D}_0$ (see ref. [46]), the level-shift parameter μ shifts only the energies of the occupied MOs. Therefore, the role of μ is to modify the difference between the energies of the occupied and virtual MOs - in particular, the HOMO–LUMO gap.

Clearly, the success of the trust region Roothaan–Hall (TRRH) method will depend on our ability to make a judicious choice of the level-shift parameter μ in Eq. (1.28). In our standard TRRH implementation, we determine μ by requiring that $\mathbf{D}(\mu)$ does not differ much from $\mathbf{D}_0$ in the sense of

Eq. (1.24), thereby ensuring a continuous and controlled development of the density matrix from the initial guess to the converged one.

1.4.1.2 The Trust Region RH Level Shift

The constraint on the change in the AO density Eq. (1.24) refers to a change which may arise not only from small changes in many MOs but also from large changes in a few MOs or even in a single MO. To obtain a high level of control, we shall require that the changes in the individual MOs are all small. Expanding the MOs φ_i^{new}, obtained by diagonalization of Eq. (1.28), in the old MOs, we obtain

$$\varphi_i^{\text{new}} = \sum_j^{\text{occ}} \left\langle \varphi_j^{\text{old}} \middle| \varphi_i^{\text{new}} \right\rangle \varphi_j^{\text{old}} + \sum_a^{\text{virt}} \left\langle \varphi_a^{\text{old}} \middle| \varphi_i^{\text{new}} \right\rangle \varphi_a^{\text{old}} , \tag{1.29}$$

where the first summation is over the occupied MOs and the second over the virtual MOs. The squared norm of the projection of φ_i^{new} onto the MO space associated with $\mathbf{D}_0$ is therefore

$$a_i^{\text{orb}} = \sum_j \left| \left\langle \varphi_j^{\text{old}} \middle| \varphi_i^{\text{new}} \right\rangle \right|^2 . \tag{1.30}$$

To ensure small individual MO changes in each iteration (to within a unitary transformation of the occupied MOs), we shall therefore require

$$a_{\text{min}}^{\text{orb}} = \min_i a_i^{\text{orb}} \geq A_{\text{min}}^{\text{orb}} , \tag{1.31}$$

where $A_{\text{min}}^{\text{orb}}$ is close to one (0.98 or 0.975 in practice). This way of controlling the changes in the density was also used by Seeger and Pople in their steepest descent method[11].

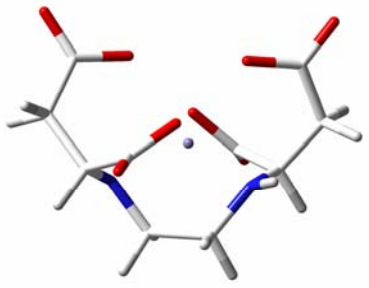


Fig. 1.3 Zn^{2+} in complex with ethylenediamine-N,N'-disuccinic acid (EDDS).

To illustrate how this scheme is used in practice, detailed information from the TRRH step in iteration 7 of a HF/6-31G and an LDA/6-31G calculation on the zinc complex depicted in Fig. 1.3 is displayed in Fig. 1.4 and Fig. 1.5, respectively. In the upper panels is illustrated how a search for $a_{\text{min}}^{\text{orb}} = A_{\text{min}}^{\text{orb}}$ determines the optimal level shift μ for the TRRH step. The TRRH energy model is more accurate for HF than for DFT (see Section 1.5.1), and consequently larger changes can be handled in the TRRH step for HF than for DFT. $A_{\text{min}}^{\text{orb}}$ is thus set to 0.975 for HF and 0.98 for DFT. In the lower panels is seen that the chosen level shifts avoid an increase in the energy which would have been the case if the Roothaan-Hall step was not level shifted (μ = 0). Notice also that an even lower energy would have been obtained by reducing the level shift, but then the restrictions on the overlap should be loosened, and this would result in

energy increase in other iterations. In short, the identification of μ from the overlap requirement $a_{\text{min}}^{\text{orb}} = A_{\text{min}}^{\text{orb}}$ appears to be a good and secure way to control the step sizes in the optimization.

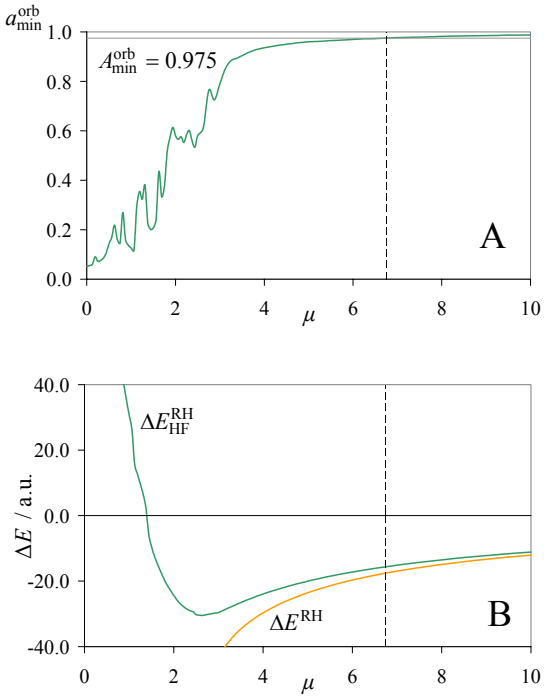


Fig. 1.4 HF/6-31G, iteration 7. (A) The overlap $a_{\text{min}}^{\text{orb}}$ and (B) the changes in the HF energy $\Delta E_{\text{HF}}^{\text{RH}}$ and in the RH energy model ΔE^{RH} as a function of the level shift μ.

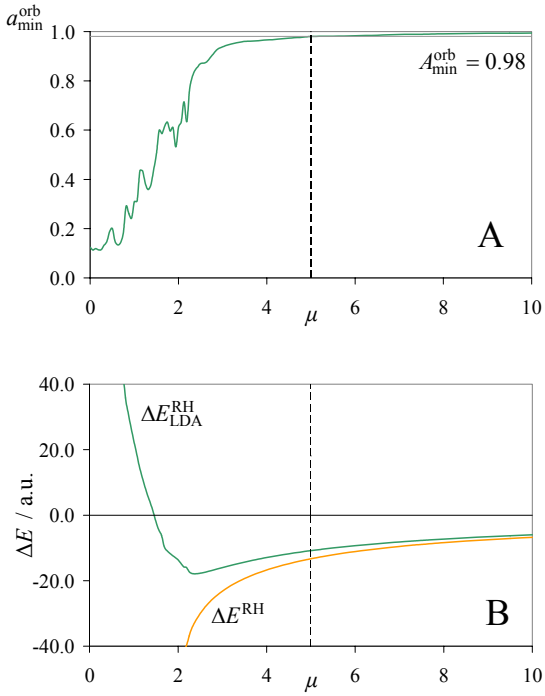


Fig. 1.5 LDA/6-31G, iteration 7. (A) The overlap $a_{\text{min}}^{\text{orb}}$ and (B) the changes in the LDA energy $\Delta E_{\text{LDA}}^{\text{RH}}$ and in the RH energy model ΔE^{RH} as a function of the level shift μ.

1.4.1.3 DIIS and Dynamically Level Shifted RH

For accelerating the SCF convergence, DIIS is a simple and in general very successful scheme. We would expect to get an even better performance and improve the stability of the scheme if DIIS was combined with a dynamically level shifted RH step like TRRH instead of the standard RH with no control of the step. To investigate how a combination of DIIS and TRRH performs, we carried out a number of DIIS-TRRH optimizations. A typical example is seen in Fig. 1.7 and an extraordinary example is seen in Fig. 1.8.

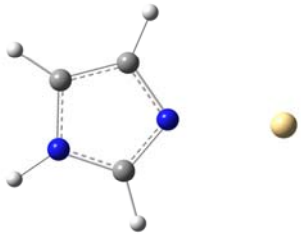


Fig. 1.6 Cd^{2+} complexed with an imidazole ring.

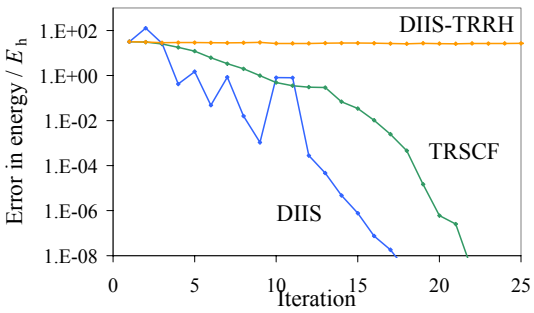


Fig. 1.7 LDA/STO-3G calculations with a H1-core start guess on the cadmium complex in Fig. 1.6.

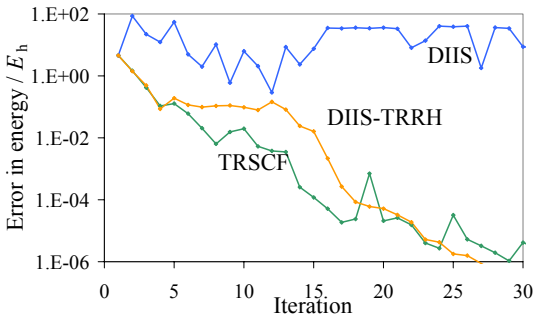


Fig. 1.8 LDA/STO-3G calculations with a Hückel start guess on the zinc complex in Fig. 1.3.

Somewhat surprisingly the calculations rarely converge with the DIIS-TRRH method. To understand this behavior, we note that, in the global region, the TRRH method typically produces gradients that do not change much, even though large changes may occur in the energy. In such cases, the DIIS method may stall, not being able to identify a good combination of density matrices. This behavior is illustrated in Table 1-1, where the gradient norm and Kohn–Sham energy of the first six iterations of the cadmium complex calculations in Fig. 1.7 are listed.

Table 1-1. The Gradient norm $\|\mathbf{g}\|=\|4(\mathbf{SDF}-\mathbf{FDS})\|$ in the first six iterations of the cadmium complex calculations of Fig. 1.7.

	DIIS		DIIS-TRRH		TRSCF	
It.	E_{KS}	$\|\mathbf{g}\|$	E_{KS}	$\|\mathbf{g}\|$	E_{KS}	$\|\mathbf{g}\|$
1	-5597.0	7.8	-5597.0	7.8	-5597.0	7.8
2	-5502.3	14.9	-5598.4	7.2	-5598.3	7.1
3	-5602.1	9.7	-5600.3	8.5	-5603.7	9.3
4	-5628.5	2.1	-5599.9	7.7	-5611.1	9.1
5	-5627.4	3.5	-5599.9	7.8	-5616.8	7.7
6	-5628.8	0.8	-5600.2	8.1	-5622.7	7.5
	conv		no conv		conv	

The TRSCF and DIIS-TRRH gradients stay almost the same during these iterations, stalling the DIIS-TRRH optimization but not the TRSCF optimization, whose energy decreases in each iteration. In the pure DIIS optimization, by contrast, the gradient changes significantly from iteration to iteration; at the same time, the energy decreases at each iteration except the second and fifth, where also the gradient norms increase. Eventually, DIIS enters the local region with its rapid rate of convergence although we note a sudden, large increase in the energy in iterations 10 and 11. However, these changes are accompanied with large increases in the gradient norm, allowing DIIS to recover safely.

In the example Fig. 1.8 standard DIIS diverges. TRSCF converges, but a minimum level shift of 0.1 is used all through the calculation. When DIIS is combined with TRRH in this case, also using a minimum level shift of 0.1, it converges as well as TRSCF. Table 1-2 contains the gradient norm and Kohn-Sham energy of the first six iterations of the calculations in Fig. 1.8.

Table 1-2. The gradient norm $\|\mathbf{g}\|=\|4(\mathbf{SDF}-\mathbf{FDS})\|$ in the first six iterations of the zinc complex calculations of Fig. 1.8.

	DIIS		DIIS-TRRH		TRSCF	
It.	E_{KS}	$\|\mathbf{g}\|$	E_{KS}	$\|\mathbf{g}\|$	E_{KS}	$\|\mathbf{g}\|$
1	-2826.95	11.6	-2826.95	11.6	-2826.95	11.6
2	-2745.49	24.0	-2830.11	3.3	-2830.06	3.4
3	-2809.38	13.6	-2831.04	1.6	-2831.11	1.5
4	-2819.16	9.7	-2831.44	0.8	-2831.42	1.1
5	-2776.74	15.4	-2831.34	1.5	-2831.40	1.5
6	-2826.55	7.0	-2831.41	1.5	-2831.47	0.9
	no conv		conv		conv	

In this case the gradient norms for the TRSCF calculation change significantly and a decrease in gradient relates directly to a decrease in the energy, where in the first example there were no direct connection between the gradient norm and the energy. The DIIS-TRRH calculation follows the same gradient behavior as TRSCF, just as in the first example, and they both converge. The DIIS gradient norm changes, but does not decrease as in the first example. There is still the connection between small gradients and low energies though, so why DIIS cannot find the proper directions in this case is not evident.

In our experience DIIS should not be used in connection with a dynamic level shift scheme like TRRH, since for all but the simplest cases DIIS-TRRH diverged if DIIS converged. We encountered, however, the example in Fig. 1.8 where DIIS does not converge and DIIS-TRRH does, but it was the exception.

1.4.1.4 Line Search TRRH

In view of the relative crudeness of the $E^{RH}(\mathbf{D})$ model, a more robust approach for choosing the level shift μ than the one presented in Section 1.4.1.2 consists of performing a line search along the path defined by μ to obtain the minimum of the energy $E_{SCF}^{RH}(\mathbf{D}(\mu))$. Strictly speaking, this optimization is not a line search but rather a univariate search. A univariate search has previously been used by Seeger and Pople[11] to stabilize convergence of the RH procedure.

For $\mu \to \infty$ Eq. (1.28) becomes equivalent to solving the eigenvalue equation

$$\mathbf{SD}_0\mathbf{SC}_{occ}^0 = \mathbf{SC}_{occ}^0\boldsymbol{\eta}, \tag{1.32}$$

where $\boldsymbol{\eta}$ has eigenvalues 1 for the set of orbitals that are occupied in $\mathbf{D}_0$ and eigenvalues 0 for the set of virtual orbitals. Eq. (1.32) thus effectively divides the molecular orbitals into a set that is occupied and a set that is unoccupied. If $\mathbf{D}_0$ is idempotent, it can be reconstructed from the occupied set of eigenvectors $\mathbf{C}_{\mathrm{occ}}^0$. If $\mathbf{D}_0$ is not idempotent, a purification of $\mathbf{D}_0$ is obtained

$$\mathbf{D}_0^{\mathrm{idem}} = \mathbf{C}_{\mathrm{occ}}^0 \left(\mathbf{C}_{\mathrm{occ}}^0 \right)^{\mathrm{T}}. \tag{1.33}$$

Since $\mathbf{F}_0$ is the gradient of $E(\mathbf{D}_0)$, the step from Eq. (1.28) corresponding to a large μ is in the steepest descent direction, and will therefore give a decrease in the Hartree-Fock energy compared to the energy at $\mathbf{D}_0$. Thus a μ exists for which the energy decreases and a line search can then find the μ leading to the largest decrease in the energy. Using the same example as in Section 1.4.1.2, Fig. 1.9 and Fig. 1.10 illustrate how the optimal μ is chosen for the line search TRRH (TRRH-LS) algorithm. A simple search in the energy change for the RH step is carried out, where the energy change is found as

$$\Delta E_{\mathrm{SCF}}^{\mathrm{RH}}(\mu) = E_{\mathrm{SCF}}(\mathbf{D}(\mu)) - E_{\mathrm{SCF}}\left(\mathbf{D}_0^{\mathrm{idem}}\right), \tag{1.34}$$

and the μ leading to the largest decrease in energy is chosen as marked on the figures.

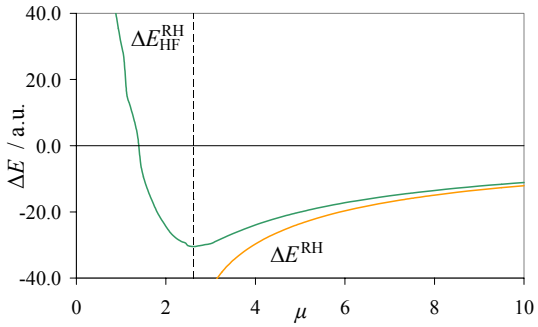


Fig. 1.9 HF/6-31G, iteration 7. The changes in the HF energy $\Delta E_{\mathrm{HF}}^{\mathrm{RH}}$ and in the RH energy model ΔE^{RH} as a function of the level shift μ.

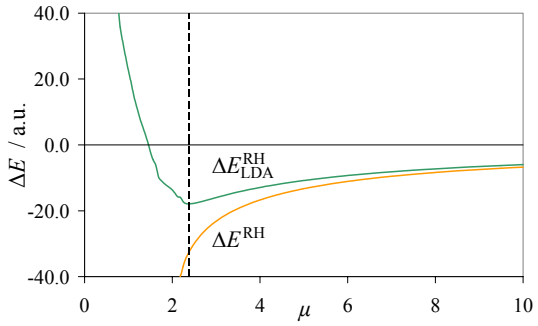


Fig. 1.10 LDA/6-31G, iteration 7. The changes in the LDA energy $\Delta E_{\mathrm{LDA}}^{\mathrm{RH}}$ and in the RH energy model ΔE^{RH} as a function of the level shift μ.

The TRRH-LS algorithm thus ensures an energy decrease in the RH step, but is of course much more expensive than the standard method, requiring the repeated construction of the Fock matrix for a single RH step. However, the first derivative $\mathrm{d}E_{\mathrm{SCF}}/\mathrm{d}\mu$ can be evaluated from the Fock matrix, and a cubic spline interpolation can thus be made from only two points on the $\Delta E_{\mathrm{SCF}}^{\mathrm{RH}}$ curve.

1.4.1.5 Optimal Level Shift without MO Information

As seen from Eq. (1.29) the individual MOs are used to find a suitable level shift in the TRRH scheme. We are very much aware that this is the most import point to improve on in our scheme. To obtain this MO information, the cubically scaling diagonalization of the Fock matrix is necessary,

and furthermore the MO coefficient matrices **C** are inherently non-sparse. Several linear or near-linear scaling alternatives to diagonalization have been suggested in the literature[18-20]. These methods could be reformulated with a dynamical level shift scheme like ours if the scheme could do without the MO information, but it is not an easy task to find a good dynamic level shift scheme with a high level of control without the knowledge of the developments in the individual MOs. The search used to find the level shift in TRRH-LS is directly applicable since it is not dependent on the MO information; the problem is only the number of Fock evaluations. The Fock evaluation is still expensive even though algorithms which make the evaluation of the Fock matrix cheaper are continually developed.

This section describes a very recently developed approach to find the optimal level shift in the TRRH step without the use of individual MOs or knowledge of the HOMO-LUMO gap. So far it has proven to be the most successful level shift scheme we have studied. The scheme is build on the assumption that the TRRH step is taken in connection with a TRDSM step (or some other density subspace minimization method). In this case it can be exploited that TRDSM is a very good energy model (see Section 1.4.2.2) and can be trusted with the responsibility to find the best direction as long as not too much new information is introduced to the density subspace in each step.

A new density, found by diagonalization of a level shifted Fock matrix or by some alternative, can be split in a part $\mathbf{D}^{\|}$ that can be described in the previous densities and a part $\mathbf{D}^{\perp}$ with new information orthogonal to the existing subspace

$$\mathbf{D}(\mu) = \mathbf{D}^{\|} + \mathbf{D}^{\perp}. \tag{1.35}$$

$\mathbf{D}^{\|}$ can be expanded in the previous densities as

$$\mathbf{D}^{\|} = \sum_{i=1}^{n} \omega_i \mathbf{D}_i \,, \tag{1.36}$$

where n is the number of previously stored densities $\mathbf{D}_i$ and the expansion coefficients ω_i are dependent on μ and determined in a least-squares manner

$$\omega_i(\mu) = \sum_{j=1}^{n} \left[\mathbf{M}^{-1} \right]_{ij} \mathrm{Tr}\, \mathbf{D}_j \mathbf{S} \mathbf{D}(\mu) \mathbf{S}, \quad M_{ij} = \mathrm{Tr}\, \mathbf{D}_i \mathbf{S} \mathbf{D}_j \mathbf{S} \,. \tag{1.37}$$

It is obvious that when $\mu \to \infty$ then $\mathbf{D}^{\perp} \to 0$ since the new density then approaches the initial density $\mathbf{D}_0$, see Eq. (1.32) and (1.33), which belongs to the set of previous densities. Thus, there is a connection between $\mathbf{D}^{\perp}$ and μ which we can exploit. If the ratio d^{orth} of the square norm $\|\mathbf{D}^{\perp}\|_{\mathrm{S}}^2$ relative to $\|\mathbf{D}\|_{\mathrm{S}}^2$ is small, only small changes to the density subspace are introduced;

$$d^{\text{orth}} = \frac{\left\| \mathbf{D}^{\perp} \right\|_{\mathbf{S}}^{2}}{\left\| \mathbf{D} \right\|_{\mathbf{S}}^{2}} = \frac{\text{Tr}\, \mathbf{D}^{\perp}\mathbf{S}\mathbf{D}^{\perp}\mathbf{S}}{\text{Tr}\, \mathbf{D}\mathbf{S}\mathbf{D}\mathbf{S}} < \delta \,, \tag{1.38}$$

where δ is some small number and $\mathbf{D}^{\perp}$ can be found as $\mathbf{D}^{\perp} = \mathbf{D} - \mathbf{D}^{\parallel}$. To illustrate how this is used in a dynamic level shift scheme, the examples from the previous sections are again seen in Fig. 1.11 and Fig. 1.12.

In the rest of the thesis the level shift scheme described in Section 1.4.1.2 will be referred to as the **C**-shift scheme since it involves the eigenvectors **C** from the diagonalization of the Fock matrix, and the level shift scheme described in this section will be referred to as the d^{orth}-shift scheme. If nothing is mentioned about the level shift scheme, the **C**-shift is implied.

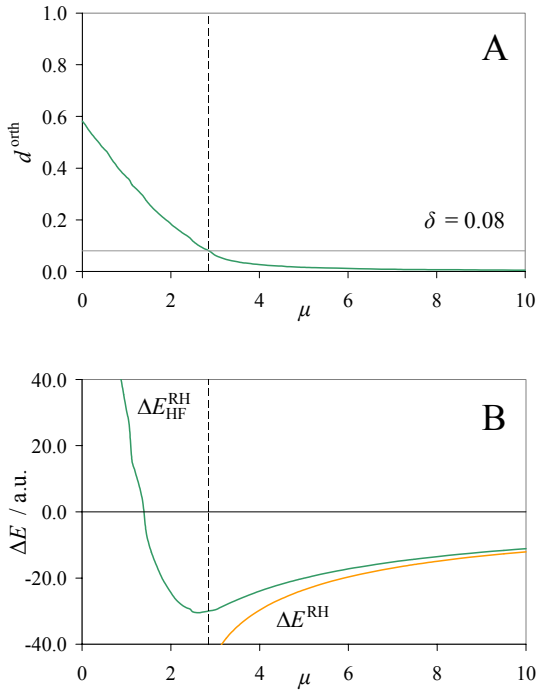


Fig. 1.11 HF/6-31G iteration 7. (A) The ratio d^{orth} and (B) the changes in the HF energy $\Delta E_{\text{HF}}^{\text{RH}}$ and in the RH energy model ΔE^{RH} as a function of the level shift μ.

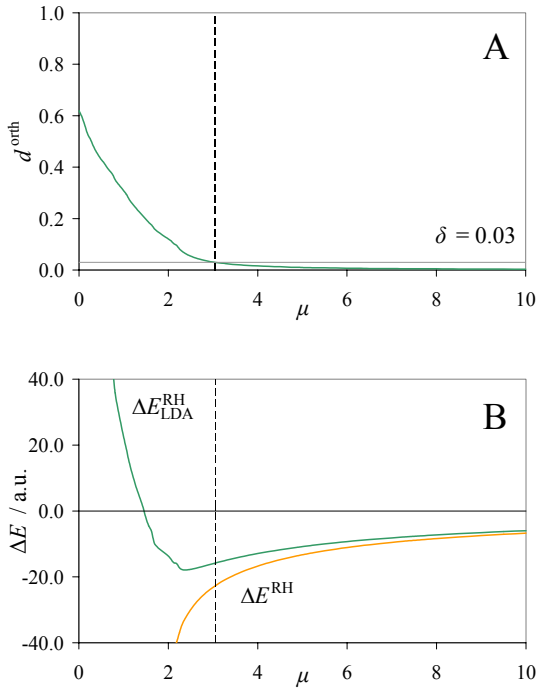


Fig. 1.12 LDA/6-31G iteration 7. (A) The ratio d^{orth} and (B) the changes in the LDA energy $\Delta E_{\text{LDA}}^{\text{RH}}$ and in the RH energy model ΔE^{RH} as a function of the level shift μ.

The upper panels now display the search made in d^{orth}, and it is clearly seen that $d^{\text{orth}} \to 0$ for $\mu \to \infty$ as expected, and increases for $\mu \to 0$. As for the **C**-shift scheme we can allow larger changes in the HF method than in DFT, and thus δ is set to 0.08 for HF and 0.03 for DFT. In the lower panels are seen that this level shift avoids an increase in the energy just as the **C**-shift scheme, but the level shift chosen here is closer to the optimal line search level shift, and thus leads to a larger decrease in the energy than was the case for the **C**-shift scheme.

In the **C**-shift scheme seen in Eq. (1.31) the changes introduced are controlled compared to the previous density, whereas in the d^{orth}-shift scheme the changes are controlled compared to the subspace of all the previous densities. This scheme is thus less restrictive than the **C**-shift scheme, but it seems that the **C**-shift scheme is too restrictive, ignoring the stability gained from the subspace information. To compare the overall effect of the two level shift schemes on the SCF convergence, calculations are given in Fig. 1.13 and Fig. 1.14, for HF and LDA, respectively. The HF calculations are on CrC with bond distance 2.00Å in the STO-3G basis and the LDA calculations are on the zinc complex seen in Fig. 1.3 in the 6-31G basis, both cases for which DIIS diverges. The starting orbitals have been obtained by diagonalization of the one-electron Hamiltonian (H1-core start guess).

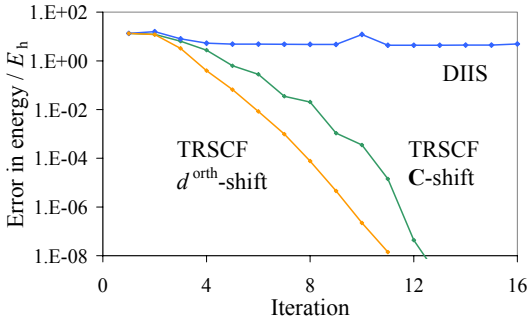


Fig. 1.13 SCF convergence for HF/STO-3G calculations on CrC.

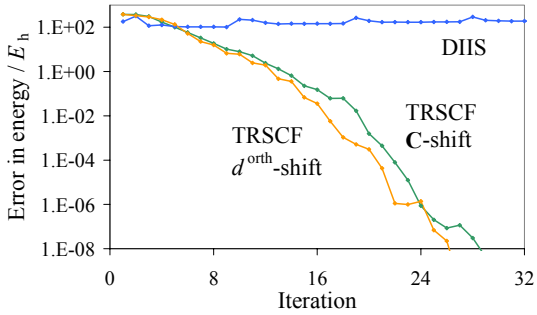


Fig. 1.14 SCF convergence for LDA/6-31G calculations on the zinc complex in Fig. 1.3.

The only difference in the "TRSCF/d^{orth}-shift" and the "TRSCF/**C**-shift" optimizations is the way the level shift is found in the TRRH step. Since DIIS diverges, the examples display the stability of the TRSCF algorithm, and the ability of the two level shifting schemes to handle problematic cases. In all examples studied so far, both problematic and simple, the d^{orth}-shift has proven as good as or better than the **C**-shift. The cost of the level shift search process is similar in the two schemes; the matrix **M** in Eq. (1.37) is updated in each iteration as a part of TRDSM and is then reused for the d^{orth}-shift scheme in TRRH.

In Table 1-3 The SCF energy change in each iteration is divided in the part of the change obtained from the RH and DSM step, respectively, and it is seen how the RH step is now allowed to accept larger changes in the density, but still in a controlled manner, thus leading to larger decreases in the energy and improved convergence.

Table 1-3. The SCF energy change for each RH and DSM step in the TRSCF calculations in Fig. 1.13.

It.	**C**-shift		d^{orth}-shift	
	$\Delta E_{\text{HF}}^{\text{RH}}$	$\Delta E_{\text{HF}}^{\text{DSM}}$	$\Delta E_{\text{HF}}^{\text{RH}}$	$\Delta E_{\text{HF}}^{\text{DSM}}$
2	-1.1768	0.0000	-1.3976	0.0000
3	-1.8964	-3.8998	-4.1319	-4.5865
4	-1.6764	-1.9603	-1.8021	-1.0448
5	-0.3655	-1.7543	-0.2103	-0.1200
6	-0.1881	-0.1624	-0.0111	-0.0463
7	-0.0932	-0.1505	-0.0036	-0.0037
8	0.0065	-0.0212	-0.0001	-0.0008
9	-0.0039	-0.0154		
10	0.0002	-0.0009		

1.4.1.6 The Trace Purification Scheme

The dynamic level shift scheme described in the previous section has no reference to the MO basis. This opens the possibility to replace the diagonalizations in the TRRH step with some alternative scheme without affecting the overall result.

There have been many suggestions as to how the diagonalization can be replaced by a linear scaling algorithm[47]. The trace purification (TP) scheme[19,48], however, is a simple and useful approach and it has thus been implemented in our SCF program in a local version of DALTON[38,49]. The trace purification scheme was originally formulated for tight binding theory by Palser and Manolopoulos[19] and later improved by Niklasson[48], and is linear scaling when formulated in an orthogonal basis. The scheme uses the trace and idempotency properties of the density to iteratively find the new density from a suitable start guess constructed from the Fock matrix.

Since the SCF optimization is formulated in the non-orthogonal AO basis to avoid the delocalized MO basis, it is necessary to transform the matrices to an orthogonal basis. This is done by a Cholesky decomposition[50] of the AO overlap matrix **S**

$$\mathbf{S} = \mathbf{L}\mathbf{L}^{\mathrm{T}}, \tag{1.39}$$

where **L** then is used to transform the Fock matrix to an orthogonal basis

$$\mathbf{F}^{\text{orth}} = \mathbf{L}^{-1}\mathbf{F}\mathbf{L}^{-\mathrm{T}}. \tag{1.40}$$

The density resulting from the trace purification scheme will also be in the orthogonal basis and should be transformed back as

$$\mathbf{D} = \mathbf{L}^{-\mathrm{T}}\mathbf{D}^{\text{orth}}\mathbf{L}^{-1}. \tag{1.41}$$

Since the AO overlap matrix does not change during the optimization, the Cholesky decomposition and the inversion of **L** can be done once and for all in the beginning of the calculation.

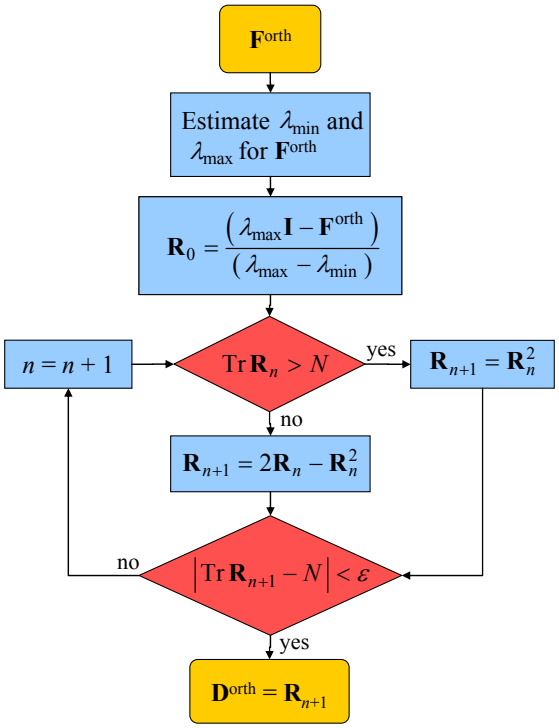


Fig. 1.15 Flow diagram for the trace purification (TP) scheme. N is the number of electrons.

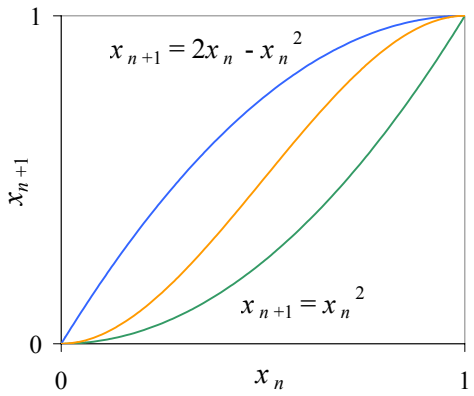


Fig. 1.16 The purifying polynomials used in the trace purification scheme. The orange line is the McWeeny purification polynomial $x_{n+1} = 3x_n^2 - 2x_n^3$.

The trace purification is carried out by the Niklasson model with second order purification polynomials, and is schematized in Fig. 1.15. The initial density guess $\mathbf{R}_0$ is obtained by normalizing the Fock matrix such that it only has eigenvalues between 0 and 1. To do this, the bounds for the Fock eigenvalues, $\lambda_{\min}$ and $\lambda_{\max}$, must be found. They can be estimated using Gerschgorin's theorem or the Lanczos algorithm for eigenvalues[51] with only a small extra computational cost. $\mathbf{R}$ is then iteratively purified, and the purification function applied in each iteration is chosen based on the trace of the matrix $\mathbf{R}$, always keeping the direction towards the correct trace condition. The purification functions are sketched in Fig. 1.16 including the McWenny purification function[8]. One of the functions used in the scheme has a stationary point for $x = 1$ and the other has a stationary point for $x = 0$; depending of the function chosen we thus go towards a larger or smaller trace. When $\mathbf{R}$ fulfils the trace and/or idempotency conditions Eq. (1.2) of the one electron density within some threshold ε, the new density $\mathbf{D}^{\text{orth}} = \mathbf{R}$ has been found and the density to use in the next TRSCF iteration can be evaluated from Eq. (1.41).

The number of purification iterations required to obtain a new density depends on the threshold ε. For the test calculations carried out so far, the threshold has been an error of 10^{-7} in the trace, and the number of iterations ranges from 30 to 70 for a single RH step, with the typical number being closer to 30 than 70. Still, it is less expensive than the diagonalization as soon as more than a couple

of thousand basis functions are needed. The scaling of the TRRH step in general and the trace purification scheme in particular is illustrated and discussed in Section 1.7.1.

1.4.2 Density Subspace Minimization

The DIIS scheme seems to have been the overall most successful of all the suggestions on how to improve SCF convergence described in Section 1.3. DIIS was the first scheme to take advantage of the information contained in the densities and Fock matrices of the previous iterations, and this made the difference.

This is also exploited in the EDIIS scheme by Kudin *et. al.*[37] in which an energy model is optimized with respect to the linear combination of previous densities. The density subspace minimization presented in this section is an improvement to EDIIS with a smaller idempotency error in the density, the correct gradient compared to SCF, and thus better convergence properties in both the local and global region of the optimization.

1.4.2.1 The Trust Region DSM Parameterization

After a sequence of Roothaan-Hall iterations, we have determined a set of density matrices $\mathbf{D}_i$ and a corresponding set of Fock matrices $\mathbf{F}_i = \mathbf{F}(\mathbf{D}_i)$. An improved density $\bar{\mathbf{D}}$ and Fock matrix $\bar{\mathbf{F}}$ should now be found as a linear combination of the previous $n + 1$ stored matrices. Taking $\mathbf{D}_0$ as the reference density matrix, the improved density matrix can be written

$$\bar{\mathbf{D}} = \mathbf{D}_0 + \sum_{i=0}^{n} c_i \mathbf{D}_i \,, \tag{1.42}$$

which, ideally, should satisfy the symmetry, trace and idempotency conditions Eq. (1.2) of a valid one-electron density matrix. Whereas the symmetry condition is trivially satisfied for any such linear combination, the trace condition holds only for combinations that satisfy the constraint

$$\sum_{i=0}^{n} c_i = 0 \,, \tag{1.43}$$

leading to a set of $n + 1$ constrained parameters c_i with $0 \leq i \leq n$. Alternatively, an unconstrained set of n parameters c_i with $1 \leq i \leq n$ can be used, with c_0 defined so that the trace condition is fulfilled:

$$c_0 = -\sum_{i=1}^{n} c_i \,. \tag{1.44}$$

In terms of these independent parameters, the density matrix $\bar{\mathbf{D}}$ becomes

$$\bar{\mathbf{D}} = \mathbf{D}_0 + \mathbf{D}_+ \,, \tag{1.45}$$

where we have introduced the notation

$$\begin{aligned} \mathbf{D}_+ &= \sum_{i=1}^{n} c_i \mathbf{D}_{i0} \\ \mathbf{D}_{i0} &= \mathbf{D}_i - \mathbf{D}_0 \,. \end{aligned} \tag{1.46}$$

Unlike the symmetry and trace conditions in Eq. (1.2), the idempotency condition is in general not fulfilled for linear combinations of $\mathbf{D}_i$. Still, for any averaged density matrix $\bar{\mathbf{D}}$ in Eq. (1.45) that does not fulfill the idempotency condition, we may generate a purified density matrix with a smaller idempotency error by the transformation[8]

$$\tilde{\mathbf{D}} = 3\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}} - 2\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}} \,. \tag{1.47}$$

Introducing the idempotency correction

$$\mathbf{D}_\delta = \tilde{\mathbf{D}} - \bar{\mathbf{D}} \,, \tag{1.48}$$

we may then write the purified averaged density matrix in the form

$$\tilde{\mathbf{D}} = \mathbf{D}_0 + \mathbf{D}_+ + \mathbf{D}_\delta \,. \tag{1.49}$$

1.4.2.2 The Trust Region DSM Energy Function

Having established a useful parameterization of the averaged density matrix Eq. (1.45) and having considered its purification Eq. (1.47), let us now consider how to determine the best set of coefficients c_i. Expanding the energy in the purified averaged density matrix, Eq. (1.49), around the reference density matrix $\mathbf{D}_0$, we obtain to second order

$$E_{\mathrm{SCF}(2)}\left(\tilde{\mathbf{D}}\right) = E_{\mathrm{SCF}}\left(\mathbf{D}_0\right) + \left(\mathbf{D}_+ + \mathbf{D}_\delta\right)^{\mathrm{T}} \mathbf{E}_0^{(1)} + \tfrac{1}{2}\left(\mathbf{D}_+ + \mathbf{D}_\delta\right)^{\mathrm{T}} \mathbf{E}_0^{(2)} \left(\mathbf{D}_+ + \mathbf{D}_\delta\right). \tag{1.50}$$

To evaluate the terms containing $\mathbf{E}_0^{(1)}$ and $\mathbf{E}_0^{(2)}$ we make the identifications

$$\mathbf{E}_0^{(1)} = 2\mathbf{F}_0 \tag{1.51}$$

$$\mathbf{E}_0^{(2)}\mathbf{D}_+ = 2\mathbf{F}_+ + \mathcal{O}\left(\mathbf{D}_+^2\right), \tag{1.52}$$

which follow from Eq. (1.4) and from the second-order Taylor expansion of $\mathbf{E}_0^{(1)}$ about $\mathbf{D}_0$. The notation Eq. (1.46) has now been generalized to the Fock matrix $\mathbf{F}_+ = \sum_{i=1}^{n} c_i \mathbf{F}_{i0}$. Ignoring the terms quadratic in $\mathbf{D}_\delta$ in Eq. (1.50) and quadratic in $\mathbf{D}_+$ in Eq. (1.52), we then obtain the DSM energy

$$E^{\mathrm{DSM}}(\mathbf{c}) = E_{\mathrm{SCF}}(\mathbf{D}_0) + 2\,\mathrm{Tr}\,\mathbf{D}_+\mathbf{F}_0 + \mathrm{Tr}\,\mathbf{D}_+\mathbf{F}_+ + 2\,\mathrm{Tr}\,\mathbf{D}_\delta\mathbf{F}_0 + 2\,\mathrm{Tr}\,\mathbf{D}_\delta\mathbf{F}_+ \,. \tag{1.53}$$

Finally, for a more compact notation, we introduce the weighted Fock matrix

$$\bar{\mathbf{F}} = \mathbf{F}_0 + \mathbf{F}_+ = \mathbf{F}_0 + \sum_{i=1}^{n} c_i \mathbf{F}_{i0} \,, \tag{1.54}$$

and find that the DSM energy may be written in the form

$$E^{\mathrm{DSM}}(\mathbf{c}) = E\left(\bar{\mathbf{D}}\right) + 2\,\mathrm{Tr}\,\mathbf{D}_\delta\bar{\mathbf{F}}, \tag{1.55}$$

where the first term is quadratic in the expansion coefficients c_i

$$E\left(\bar{\mathbf{D}}\right) = E_{\mathrm{SCF}}(\mathbf{D}_0) + 2\,\mathrm{Tr}\,\mathbf{D}_+\mathbf{F}_0 + \mathrm{Tr}\,\mathbf{D}_+\mathbf{F}_+, \tag{1.56}$$

and the second, idempotency-correction term is quartic in these coefficients:

$$2\,\mathrm{Tr}\,\mathbf{D}_\delta\bar{\mathbf{F}} = \mathrm{Tr}\left(6\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}} - 4\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}} - 2\bar{\mathbf{D}}\right)\bar{\mathbf{F}}. \tag{1.57}$$

The derivatives of $E^{\mathrm{DSM}}(\mathbf{c})$ are straightforwardly obtained by inserting the expansions of $\bar{\mathbf{F}}$ and $\bar{\mathbf{D}}$, using the independent parameter representation. The expressions are given in **Error! Reference source not found.**.

The energy function $E^{\mathrm{DSM}}(\mathbf{c})$ in Eq. (1.55) provides an excellent approximation to the exact SCF energy $E_{\mathrm{SCF}}(\mathbf{c})$ about $\mathbf{D}_0$, with an error quadratic in $\mathbf{D}_\delta$ (see Section 1.5.2). The EDIIS energy model corresponds to the first term $E(\bar{\mathbf{D}})$ in Eq. (1.55) and has thus an error linear in $\mathbf{D}_\delta$.

1.4.2.3 The Trust Region DSM Minimization

The DSM energy, Eq. (1.55), is minimized with respect to the independent parameters c_i with $1 \le i \le n$. The vector containing the parameters is initialized to zero $\mathbf{c}^{(0)} = \mathbf{0}$ such that $\bar{\mathbf{D}} = \mathbf{D}_0$, where $\mathbf{D}_0$ is chosen as the density matrix with the lowest energy $E_{\mathrm{SCF}}(\mathbf{D}_i)$, usually the one from the latest TRRH step. The minimization is then carried out by the trust region method[52], taking a number of steps from the initial parameters $\mathbf{c}^{(0)}$ to the final optimized parameters $\mathbf{c}^*$ as illustrated in Fig. 1.17.

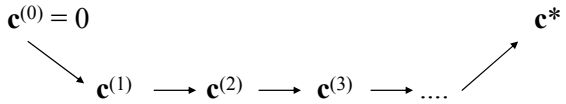


Fig. 1.17 Steps in the trust region minimization of the DSM energy.

We thus consider in each step the second-order Taylor expansion of the DSM energy in Eq. (1.55). Introducing the step vector

$$\Delta\mathbf{c} = \mathbf{c}^{(i+1)} - \mathbf{c}^{(i)}, \tag{1.58}$$

we obtain

$$E^{\mathrm{DSM}}_{(2)}\left(\mathbf{c}^{(i)} + \Delta\mathbf{c}\right) = E_0 + \Delta\mathbf{c}^{\mathrm{T}}\mathbf{g} + \tfrac{1}{2}\Delta\mathbf{c}^{\mathrm{T}}\mathbf{H}\Delta\mathbf{c}, \tag{1.59}$$

where the energy, gradient, and Hessian at the expansion point are given by

$$E_0 = E^{\mathrm{DSM}}\left(\mathbf{c}^{(i)}\right), \quad \mathbf{g} = \left.\frac{\partial E^{\mathrm{DSM}}(\mathbf{c})}{\partial\mathbf{c}}\right|_{\mathbf{c}=\mathbf{c}^{(i)}}, \quad \mathbf{H} = \left.\frac{\partial^2 E^{\mathrm{DSM}}(\mathbf{c})}{\partial\mathbf{c}^2}\right|_{\mathbf{c}=\mathbf{c}^{(i)}}. \tag{1.60}$$

We then introduce a trust region of radius h for $E_{(2)}^{\mathrm{DSM}}\left(\mathbf{c}^{(i)}+\mathbf{\Delta c}\right)$ and require that steps are always taken inside or to the boundary of this region. To determine a step to the boundary, we restrict the step to have the length h in the S metric norm $\tilde{\mathbf{M}}$

$$\|\mathbf{\Delta c}\|_S^2 = \sum_{ij=1}^{n} \Delta c_i \tilde{M}_{ij} \Delta c_j = h^2 \,. \tag{1.61}$$

In the unconstrained formulation defined by Eq. (1.44), the metric $\mathbf{M}$ of Eq. (1.37), is found as

$$\tilde{M}_{ij} = \mathrm{Tr}\,\mathbf{D}_i\mathbf{S}\mathbf{D}_j\mathbf{S} - \mathrm{Tr}\,\mathbf{D}_i\mathbf{S}\mathbf{D}_0\mathbf{S} - \mathrm{Tr}\,\mathbf{D}_0\mathbf{S}\mathbf{D}_j\mathbf{S} + \mathrm{Tr}\,\mathbf{D}_0\mathbf{S}\mathbf{D}_0\mathbf{S}\,, \quad i,j \neq 0\,, \tag{1.62}$$

Introducing the undetermined multiplier ν for the step-size constraint, we arrive at the following Lagrangian for minimization on the boundary of the trust region:

$$L\left(\mathbf{\Delta c},\nu\right) = E_0 + \mathbf{\Delta c}^{\mathrm{T}}\mathbf{g} + \tfrac{1}{2}\mathbf{\Delta c}^{\mathrm{T}}\mathbf{H}\mathbf{\Delta c} - \tfrac{1}{2}\nu\left(\mathbf{\Delta c}^{\mathrm{T}}\tilde{\mathbf{M}}\mathbf{\Delta c} - h^2\right). \tag{1.63}$$

Differentiating this Lagrangian and setting the derivatives equal to zero, we obtain the equations

$$\frac{\partial L}{\partial \mathbf{\Delta c}} = \mathbf{g} + \mathbf{H}\mathbf{\Delta c} - \nu\tilde{\mathbf{M}}\mathbf{\Delta c} = 0 \tag{1.64}$$

$$\frac{\partial L}{\partial \nu} = -\tfrac{1}{2}\left(\mathbf{\Delta c}^{\mathrm{T}}\tilde{\mathbf{M}}\mathbf{\Delta c} - h^2\right) = 0\,. \tag{1.65}$$

The optimization of the Lagrangian thus corresponds to the solution of the following set of linear equations:

$$\left(\mathbf{H} - \nu\tilde{\mathbf{M}}\right)\mathbf{\Delta c} = -\mathbf{g}\,, \tag{1.66}$$

where the multiplier ν is iteratively adjusted until the step is to the boundary of the trust region Eq. (1.65). The step length restriction may be lifted by setting $\nu = 0$ as needed for steps inside the trust region.

To illustrate how the level shift parameter ν in Eq. (1.66) is determined, we consider in Fig. 1.18 and Fig. 1.19 the third and fourth DSM step respectively, in iteration five of the HF/STO-3G calculation on CrC seen in Fig. 1.13. The step length $\|\mathbf{\Delta c}\|_S$ is plotted as a function of ν. The plots consist of branches between asymptotes where ν makes the matrix on the left hand side of Eq. (1.66) singular. This happens whenever ν equals one of the Hessian eigenvalues. The lowest eigenvalue ω_1 of the Hessian $\mathbf{H}$ is found, and the level shift parameter is chosen in the interval $-\infty < \nu < \min(0,\omega_1)$. The proper value is found where the step length function crosses the line representing the trust radius h, as marked in Fig. 1.18. If the step that minimizes $E_{(2)}^{\mathrm{DSM}}$ is inside the trust region, $\nu = 0$ is chosen as is the case in Fig. 1.19. The trust region is updated during the iterative procedure and therefore h is different in the two steps.

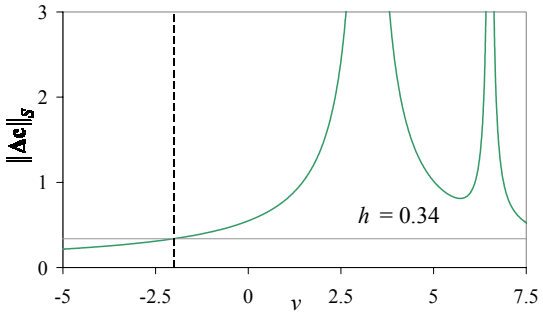


Fig. 1.18 The step length as a function of the multiplier ν in the third DSM step.

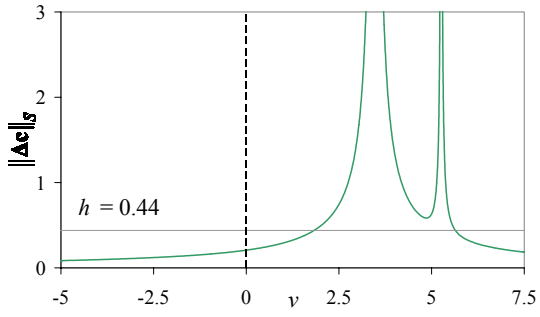


Fig. 1.19 The step length as a function of the multiplier ν in the fourth DSM step.

Each of the trust region steps require the construction of the gradient **g** and the Hessian **H** in the density subspace, and the solution of the level shifted Newton equations Eq. (1.66). Since E^{DSM} is a local model of the true energy function E_{SCF}, it resembles E_{SCF} only in a small region about the initial point $\mathbf{c}^{(0)}$. The DSM iterations are therefore terminated if the total step length after p iterations $\|\mathbf{c}^{(p)} - \mathbf{c}^{(0)}\|_S$ exceeds some preset value k. If a minimum of E^{DSM} is found inside the trust region $\|\mathbf{c}^{(p)} - \mathbf{c}^{(0)}\|_S < k$, then the step $\|\mathbf{c}^* - \mathbf{c}^{(0)}\|_S$ to the minimum is taken and the iterations are terminated. This is the typical situation.

When the trust region minimization has terminated, an improved density matrix $\tilde{\mathbf{D}}$ can be constructed. However, to avoid the expensive calculation of the Fock matrix from $\tilde{\mathbf{D}}$ we use instead the averaged density matrix from eq. (1.45) and exploit that the Fock matrix is linear in the density for Hartree-Fock such that $\mathbf{F}(\bar{\mathbf{D}})$ is simply the averaged Fock matrix of Eq. (1.54). For DFT this is an approximation, but typically insignificant improvements are obtained by evaluating the correct Kohn-Sham matrix. The improved Fock matrix and density matrix then enters the TRRH step as $\mathbf{F}_0$ and $\mathbf{D}_0$, respectively.

By construction $E^{\mathrm{DSM}}(\mathbf{c})$ is lowered at each iteration of the trust region minimization. Since E^{DSM} is a local model to the true energy E_{SCF}, the lowering of E^{DSM} will also lead to a lowering of E_{SCF} provided the total step is sufficiently short and thus stays in the local region.

1.4.2.4 Line Search TRDSM

As in the TRRH step, the averaged density matrix $\bar{\mathbf{D}}$ may also be determined by a line search and we denote this line search algorithm TRDSM-LS. Here, the line search is made in the direction defined by the first step $\mathbf{c}^{(1)}$ of the TRDSM algorithm—that is, the step at the expansion point $\mathbf{D}_0$. As in the TRRH step, such a line search is guaranteed to reduce the energy. The first step is scaled by a parameter α,

$$\mathbf{\Delta c}_{\text{tot}} = \alpha \cdot \mathbf{c}^{(1)} \tag{1.67}$$

and a search is made in $\Delta E_{\text{SCF}}^{\text{DSM}}$ to find the step $\mathbf{\Delta c}_{\text{tot}}$ that leads to the largest decrease in energy. $E_{\text{SCF}}(\alpha)$ is found by evaluating the averaged density of Eq. (1.45) for the coefficients ($\mathbf{c}_0 + \mathbf{\Delta c}_{\text{tot}}$), purifying it as in Eq. (1.32)–(1.33) and inserting it in the energy expression of Eq. (1.1). Then $\Delta E_{\text{SCF}}^{\text{DSM}}(\alpha)$ can be found as

$$\Delta E_{\text{SCF}}^{\text{DSM}}(\alpha) = E_{\text{SCF}}(\alpha) - E_{\text{SCF}}(\mathbf{D}_0). \tag{1.68}$$

Fig. 1.20 and Fig. 1.21 illustrate the search in α, again for iteration seven of the HF and LDA calculations on the zinc complex in Fig. 1.3. For $\alpha = 0$, no step is taken and hence no energy decrease is seen. For the marked choice of α, the optimal step length is obtained.

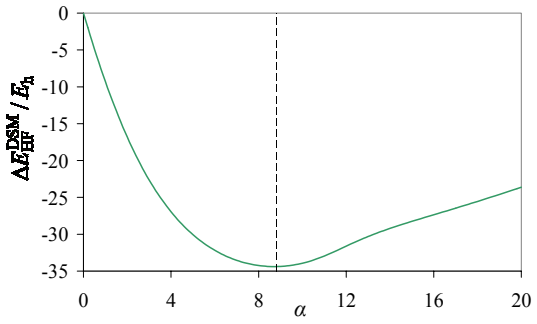


Fig. 1.20 Decrease in HF energy as a function of the step length α.

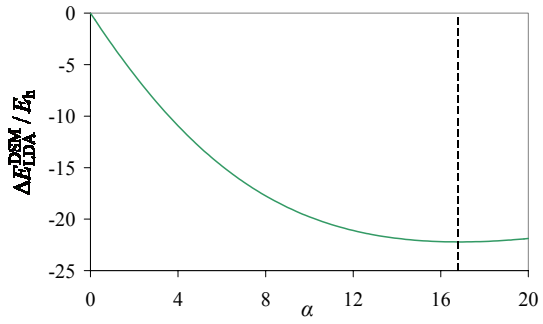


Fig. 1.21 Decrease in LDA energy as a function of the step length α.

1.4.2.5 The Missing Term

In the construction of the TRDSM energy model Eq. (1.55), the term of second order in the idempotency correction $\mathbf{D}_\delta$ was neglected from Eq. (1.50), since this term required a new Fock evaluation $\mathbf{F}(\mathbf{D}_\delta)$, which would increase the expenses of the scheme considerably. This section will be concerned with this neglected term and how a part of it can be described without the evaluation of a new Fock matrix, leading to an improved energy model for TRDSM at no considerable extra cost. The actual effect of this improvement to the energy model will then be discussed through a case study. This section will only be concerned with Hartree-Fock theory and examples, but it might equally well be done for DFT even though the improvement should be less significant since for DFT, also terms of order $\|\mathbf{D}_+\|^3$ are neglected. These are of the same size as the neglected term quadratic in $\mathbf{D}_\delta$. In Section 1.5.2 these errors are discussed.

Since the only neglect in the DSM energy model Eq. (1.55) for Hartree-Fock is the term quadratic in $\mathbf{D}_\delta$, and since the only term quadratic in the density is Tr$\mathbf{DG}(\mathbf{D})$, the HF energy for the density $\tilde{\mathbf{D}}$ can be written as

$$E_{\mathrm{HF}}\left(\tilde{\mathbf{D}}\right) = E\left(\bar{\mathbf{D}}\right) + 2\,\mathrm{Tr}\,\mathbf{D}_\delta\bar{\mathbf{F}} + \mathrm{Tr}\,\mathbf{D}_\delta\mathbf{G}\left(\mathbf{D}_\delta\right), \tag{1.69}$$

where $E\left(\bar{\mathbf{D}}\right)$ is seen in Eq. (1.56). Even though a new Fock matrix $\mathbf{h} + \mathbf{G}(\mathbf{D}_\delta)$ should be evaluated to describe the last term exactly, a part of the term can be described in the subspace of the previous densities.

As exploited in the level-shift scheme Section 1.4.1.5, a density or density difference, in this case $\mathbf{D}_\delta$, can be divided in a part that can be described in the subspace of the previous densities $\mathbf{D}_\delta^{\parallel}$ and an unknown part orthogonal to the space $\mathbf{D}_\delta^{\perp}$

$$\mathbf{D}_\delta = \mathbf{D}_\delta^{\parallel} + \mathbf{D}_\delta^{\perp}. \tag{1.70}$$

$\mathbf{D}_\delta^{\parallel}$ is expanded in the previous densities $\mathbf{D}_i$ as

$$\mathbf{D}_\delta^{\parallel} = \sum_{i=0}^{n}\omega_i\mathbf{D}_i\,, \tag{1.71}$$

where the expansion coefficients ω_i are determined in a least-squares manner

$$\omega_i = \sum_{j=0}^{n}\left[\mathbf{M}^{-1}\right]_{ij}\mathrm{Tr}\,\mathbf{D}_j\mathbf{S}\mathbf{D}_\delta\mathbf{S}, \quad M_{ij} = \mathrm{Tr}\,\mathbf{D}_i\mathbf{S}\mathbf{D}_j\mathbf{S}\,. \tag{1.72}$$

Inserting Eq. (1.70) for $\mathbf{D}_\delta$ in Eq. (1.69), an improved DSM energy model can be written

$$E_{\mathrm{imp}}^{\mathrm{DSM}}\left(\mathbf{c}\right) = E\left(\bar{\mathbf{D}}\right) + 2\,\mathrm{Tr}\,\mathbf{D}_\delta\bar{\mathbf{F}} + \mathrm{Tr}\left(2\mathbf{D}_\delta - \mathbf{D}_\delta^{\parallel}\right)\mathbf{G}\left(\mathbf{D}_\delta^{\parallel}\right), \tag{1.73}$$

where only previous density and Fock matrices enter. The relation

$$\mathrm{Tr}\,\mathbf{A}\mathbf{G}(\mathbf{B}) = \mathrm{Tr}\,\mathbf{B}\mathbf{G}(\mathbf{A}) \tag{1.74}$$

for symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ is used and the term $\mathrm{Tr}\,\mathbf{D}_\delta^{\perp}\mathbf{G}\left(\mathbf{D}_\delta^{\perp}\right)$ is neglected. A second order Taylor expansion of the improved DSM energy can then be made as in Eq. (1.59) and a trust region minimization carried out.

To study the improvement to the energy function, two TRSCF calculations are carried out on the cadmium complex seen in Fig. 1.6 in the STO-3G basis and with a H1-core start guess. The convergence profiles of the calculations are displayed in Fig. 1.22, the one denoted “Improved TRDSM” is a TRSCF calculation just as the one denoted “TRSCF” with the only difference that the improved energy model in Eq. (1.73) is used for TRDSM instead of the one in Eq. (1.55). To illustrate the impact of the improvement in a single TRDSM step, a line search like the one in Fig. 1.20 is made in iteration 7 of the same TRSCF calculation as in Fig. 1.22. Apart from displaying the change in SCF energy as a function of the step length α, also the DSM energy of Eq. (1.55) and the improved DSM energy of Eq. (1.73) are evaluated for the different choices of α, and their energy changes found as well.

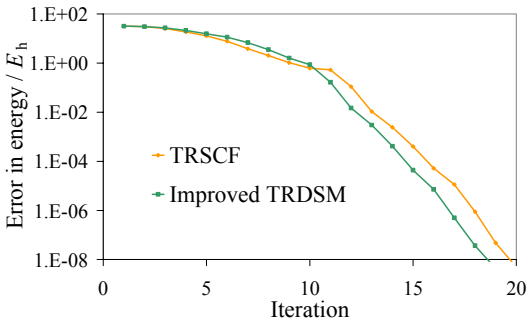


Fig. 1.22 Convergence for the cadmium complex in Fig. 1.6, both for TRSCF with no improvements, and for TRSCF where $E_{\mathrm{imp}}^{\mathrm{DSM}}$ is used in TRDSM.

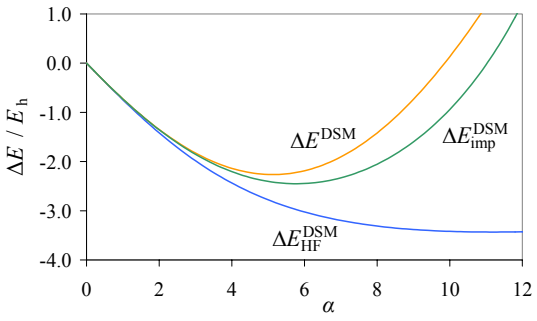


Fig. 1.23 TRDSM line search for iteration 7 in the TRSCF optimization Fig. 1.22. For different α in Eq. (1.67), the changes in $E_{\mathrm{HF}}^{\mathrm{DSM}}$, E^{DSM} and $E_{\mathrm{imp}}^{\mathrm{DSM}}$ compared to $E_{\mathrm{HF}}(\mathbf{D}_0)$ are found.

It is seen in Fig. 1.23 that the improved DSM energy describes the HF energy better than the standard DSM energy does, just as expected. As the step moves away from the expansion point, the part of the energy which cannot be described in the old densities grows and both the DSM energy models become poor.

The improvements presented in this section add complexity to the TRDSM algorithm, even though the computational cost is not significant. As seen in Fig. 1.22 and Fig. 1.23, the improvements to the TRSCF calculation are minor. The overall gain does not justify the extra complexity added to the TRDSM algorithm.

1.4.3 Energy Minimization Exploiting the Density Subspace

Section 1.3.1 describes how different approaches have been taken to avoid the diagonalization in the Roothaan-Hall step. Replacing the standard diagonalization of the Fock matrix can be done for the purpose of improving either the convergence properties or the scaling of the algorithm or for both reasons. With the purpose of improving both, a newly developed scheme is presented in this section, in which an energy minimization replaces the standard diagonalization in the SCF optimization.

When the RH energy model is minimized, the density subspace information used with great success in TRDSM is ignored. The novel idea is thus to exploit the valuable information saved in the density subspace of the previous densities to construct an improved RH energy model and minimize this model instead of the RH model. This makes the TRDSM step redundant since a density subspace minimization now is included in the RH energy model minimization.

The Hessian update methods[40,53], in which an approximate Hessian is updated in each iteration and an approximate Newton step is taken, exploit some of the same ideas, but they are all based on

approximate second order energy expansions in the orbital rotation parameters and therefore do not include the third and higher order terms included in the RH energy.

In the following subsections the improved RH energy model and its minimization will be described. The SCF convergence of a test case is then displayed, in which the new energy minimization approach is compared to standard DIIS and the TRSCF schemes. As the scheme has not yet been extended to DFT, this section will only consider HF theory and calculations.

1.4.3.1 The Augmented RH Energy model

If the Hartree-Fock energy, Eq. (1.1), is expanded through second order around some reference density $\mathbf{D}_0$

$$E_{\mathrm{HF}}(\mathbf{D}) = E_{\mathrm{HF}}(\mathbf{D}_0) + 2\,\mathrm{Tr}\,\mathbf{F}(\mathbf{D}_0)(\mathbf{D}-\mathbf{D}_0) + \mathrm{Tr}(\mathbf{D}-\mathbf{D}_0)\mathbf{G}(\mathbf{D}-\mathbf{D}_0), \tag{1.75}$$

the first two terms are recognized as $E^{\mathrm{RH}}(\mathbf{D})$ from Eq. (1.22) plus the terms of zeroth order $E_{\mathrm{HF}}(\mathbf{D}_0)$ and - $E^{\mathrm{RH}}(\mathbf{D}_0)$

$$E_{\mathrm{HF}}(\mathbf{D}) = E^{\mathrm{RH}}(\mathbf{D}) + \left(E_{\mathrm{HF}}(\mathbf{D}_0) - E^{\mathrm{RH}}(\mathbf{D}_0)\right) + \mathrm{Tr}(\mathbf{D}-\mathbf{D}_0)\mathbf{G}(\mathbf{D}-\mathbf{D}_0). \tag{1.76}$$

In a standard RH step, the energy function to minimize is the RH energy, neglecting the last term which contains the Hessian information, because it is too expensive to evaluate. Since Hessian information is very valuable to an optimization, the scheme presented in this section will replace the diagonalization in the RH step by an energy minimization of an augmented RH (ARH) energy model, where as much Hessian information as possible is included without directly evaluating new Fock matrices. This is done by exploiting the information contained in the density and Fock matrices of the previous iterations.

As previously exploited, a density or density difference, in this case $\mathbf{\Delta} = \mathbf{D} - \mathbf{D}_0$, can be split in a part that can be described in the subspace of the $n + 1$ previous densities $\mathbf{\Delta}^{\|}$ and an unknown part orthogonal to the space $\mathbf{\Delta}^{\perp}$

$$\mathbf{D} - \mathbf{D}_0 = \mathbf{\Delta} = \mathbf{\Delta}^{\|} + \mathbf{\Delta}^{\perp}. \tag{1.77}$$

$\mathbf{\Delta}^{\|}$ is expanded in the previous densities $\mathbf{D}_i$ as

$$\mathbf{\Delta}^{\|} = \sum_{i=0}^{n} \omega_i \mathbf{D}_i\,, \tag{1.78}$$

where n is the number of previously stored densities and the expansion coefficients ω_i are determined in a least-squares manner

$$\omega_i = \sum_{j=0}^{n} \left[\mathbf{M}^{-1}\right]_{ij} \mathrm{Tr}\,\mathbf{D}_j\mathbf{S}\mathbf{\Delta}\mathbf{S}, \quad M_{ij} = \mathrm{Tr}\,\mathbf{D}_i\mathbf{S}\mathbf{D}_j\mathbf{S}\,. \tag{1.79}$$

Inserting Eq. (1.77) in the last term of Eq. (1.76) and neglecting the term $\mathrm{Tr}\,\boldsymbol{\Delta}^{\perp}\mathbf{G}\left(\boldsymbol{\Delta}^{\perp}\right)$, the augmented Roothaan-Hall energy model can be written as

$$E^{\mathrm{ARH}}(\mathbf{D}) = E^{\mathrm{RH}}(\mathbf{D}) + \left(E_{\mathrm{HF}}(\mathbf{D}_0) - E^{\mathrm{RH}}(\mathbf{D}_0)\right) + \mathrm{Tr}\left(2\boldsymbol{\Delta} - \boldsymbol{\Delta}^{\parallel}\right)\mathbf{G}\left(\boldsymbol{\Delta}^{\parallel}\right), \tag{1.80}$$

where $\mathbf{G}\left(\boldsymbol{\Delta}^{\parallel}\right)$ is evaluated as a linear combination of previous Fock matrices

$$\mathbf{G}\left(\boldsymbol{\Delta}^{\parallel}\right) = \sum_{i=1}^{n} \omega_i \mathbf{G}(\mathbf{D}_i) = \sum_{i=1}^{n} \omega_i \left(\mathbf{F}(\mathbf{D}_i) - \mathbf{h}\right). \tag{1.81}$$

The energy model E^{ARH} has no intrinsic restrictions with respect to how different the densities spanning the subspace are allowed to be, and this is one of the benefits compared to the TRSCF scheme. For the TRDSM energy model, the purification implicit in the DSM energy makes no sense if the densities are too different, in particular if they have different electron configurations. In ARH, configuration shifts can be handled without problems, and whereas old, obsolete densities pollute the DSM energy model, they simply disappear from the ARH energy model, since their weights ω_i diminish.

We expect a faster convergence rate for ARH compared to TRSCF, mainly because the RH and DSM steps are merged to an energy model with correct gradient (not just in the subspace) and an approximate Hessian, which is improved in each iteration using the information from the previous density and Fock matrices.

1.4.3.2 The Augmented RH Optimization

The density for which the ARH energy model should be optimized can be expanded in the anti-symmetric matrix $\mathbf{X}$

$$\mathbf{D}(\mathbf{X}) = \exp(-\mathbf{XS})\,\mathbf{D}_0^{(i)} \exp(\mathbf{SX}) = \mathbf{D}_0^{(i)} + \left[\mathbf{D}_0^{(i)}, \mathbf{X}\right]_{\mathrm{S}} + \tfrac{1}{2}\left[\left[\mathbf{D}_0^{(i)}, \mathbf{X}\right]_{\mathrm{S}}, \mathbf{X}\right]_{\mathrm{S}} + \cdots, \tag{1.82}$$

where $\mathbf{D}_0^{(i)}$ is the reference density from which the step $\mathbf{X}$ is taken. Optimizing the ARH energy is thus a nonlinear problem and an iterative scheme should be applied.

A Newton-Raphson (NR) optimization of the ARH energy is therefore carried out, and the steps are found minimizing a second order approximation of the ARH energy $E_{(2)}^{\mathrm{ARH}}$ by the preconditioned conjugate gradient (PCG) method. The second order approximation of the ARH energy, where the constant terms are excluded, can be written as

$$
\begin{aligned}
E_{(2)}^{\mathrm{ARH}}(\mathbf{X}) &= 2\,\mathrm{Tr}\,\mathbf{F}_0\left[\mathbf{D}_0^{(i)},\mathbf{X}\right]_{\mathrm{S}} + \mathrm{Tr}\,\mathbf{F}_0\left[\left[\mathbf{D}_0^{(i)},\mathbf{X}\right]_{\mathrm{S}},\mathbf{X}\right]_{\mathrm{S}} \\
&+2\,\mathrm{Tr}\left(\mathbf{D}_0^{(i)}-\mathbf{D}_0\right)\sum_{i=1}^{n}\left(\omega_i^{(1)}+\omega_i^{(2)}\right)\mathbf{G}(\mathbf{D}_i) \\
&+2\,\mathrm{Tr}\left[\mathbf{D}_0^{(i)},\mathbf{X}\right]_{\mathrm{S}}\sum_{i=1}^{n}\left(\omega_i^{(0)}+\omega_i^{(1)}\right)\mathbf{G}(\mathbf{D}_i) + \mathrm{Tr}\left[\left[\mathbf{D}_0^{(i)},\mathbf{X}\right]_{\mathrm{S}},\mathbf{X}\right]_{\mathrm{S}}\sum_{i=1}^{n}\omega_i^{(0)}\mathbf{G}(\mathbf{D}_i) \\
&-\mathrm{Tr}\sum_{i,j=1}^{n}\left[2\omega_j^{(0)}\left(\omega_i^{(1)}+\omega_i^{(2)}\right)+\omega_i^{(1)}\omega_j^{(1)}\right]\mathbf{D}_i\mathbf{G}(\mathbf{D}_j),
\end{aligned}
\tag{1.83}
$$

where

$$
\begin{aligned}
\omega_i^{(0)} &= \sum_{j=1}^{n}\left[\mathbf{M}^{-1}\right]_{ij}\mathrm{Tr}\left(\mathbf{D}_j\mathbf{S}\mathbf{D}_0^{(i)}\mathbf{S}\right) \\
\omega_i^{(1)} &= \sum_{j=1}^{n}\left[\mathbf{M}^{-1}\right]_{ij}\mathrm{Tr}\left(\mathbf{D}_j\mathbf{S}\left[\mathbf{D}_0^{(i)},\mathbf{X}\right]_{\mathrm{S}}\mathbf{S}\right) \\
\omega_i^{(2)} &= \tfrac{1}{2}\sum_{j=1}^{n}\left[\mathbf{M}^{-1}\right]_{ij}\mathrm{Tr}\left(\mathbf{D}_j\mathbf{S}\left[\left[\mathbf{D}_0^{(i)},\mathbf{X}\right]_{\mathrm{S}},\mathbf{X}\right]_{\mathrm{S}}\mathbf{S}\right).
\end{aligned}
\tag{1.84}
$$

If the summations are put in the most favorable way, the number of matrix multiplications is limited and independent of subspace size. Only the update of the metric **M** takes a number of matrix multiplications linearly in the subspace size.

From the derivative $\frac{\partial E_{(2)}^{\mathrm{ARH}}}{\partial \mathbf{X}}$, the problem to be solved by PCG is set up for the current reference density $\mathbf{D}_0^{(i)}$ where i denotes the Newton-Raphson step number. Through the whole NR optimization $\mathbf{D}_0$ and $\mathbf{F}_0$ are the density and Fock matrices from the previous SCF iteration. The NR step **X** found by PCG is used to evaluate a new density from Eq. (1.82) and if the new density is similar to the previous one, the Newton-Raphson optimization has converged, if not, the density is used as reference density $\mathbf{D}_0^{(i)}$ in the next step.

The final density matrix resulting from the NR optimization is then used to evaluate a new Fock matrix, and so the SCF iterative procedure is established. The SCF scheme for the described algorithm is illustrated in Fig. 1.24.

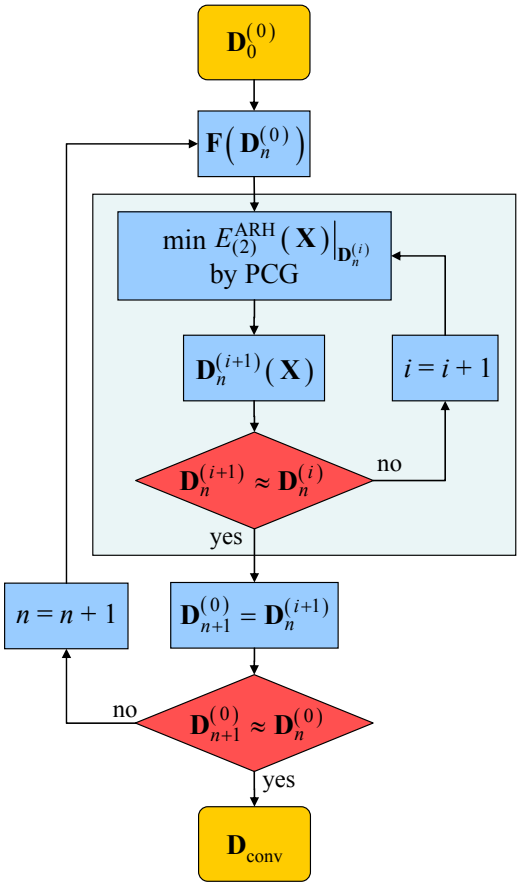


Fig. 1.24 Flow diagram of the SCF optimization with the diagonalization of the Fock matrix replaced by a minimization of the ARH energy. The light blue box embraces the Newton-Raphson optimization of E^{ARH}.

1.4.3.3 Applications

SCF calculations have been carried out using the ARH scheme. In Fig. 1.25 the convergence of HF/STO-3G calculations on CrC with 2.00Å bond distance are displayed. Results are given for the augmented RH scheme, DIIS and TRSCF with the **C**-shift and d^{orth}-shift schemes, respectively. For the first iterations in the ARH optimization a limit is put on the $\|\mathbf{X}\|_S$ norm to avoid changes in the densities which go beyond the region that is well described by the energy model.

The ARH scheme is clearly superior for this test case, even with the convergence improvements for TRSCF obtained with the d^{orth}-shift scheme; ARH is almost an iteration in front of 'TRSCF/d^{orth}-shift' in the local region. The standard DIIS approach does not converge at all for this case.

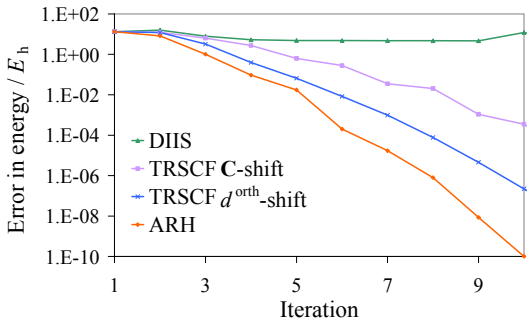


Fig. 1.25 HF/STO-3G calculations on CrC using different approaches.

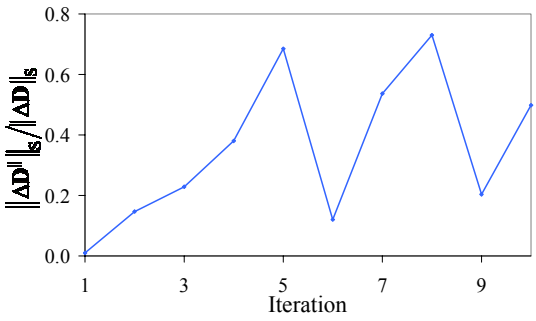


Fig. 1.26 Details from the ARH optimization in Fig. 1.25: The part of the density change which can be described in the subspace of the previous densities.

To illustrate how information gradually is obtained from the previous densities in ARH, the part of the density change $\Delta\mathbf{D} = \mathbf{D}_{n+1} - \mathbf{D}_n$ in each iteration that can be described in the previous densities $\Delta\mathbf{D}^{\|}$ is found as in Eq. (1.78)-(1.79), and the ratio $\|\Delta\mathbf{D}^{\|}\|_S / \|\Delta\mathbf{D}\|_S$ is depicted in Fig. 1.26. It is seen how the description of $\Delta\mathbf{D}$ improves during the first five iterations until a significant part of the Hessian is described, then a qualified step is taken to another region, and the new density is therefore not well described in the previous densities. This step is followed by a significant decrease in SCF energy of two orders of magnitude. The same pattern is repeated after two additional iterations.

Even though only preliminary results are given in this section, the ARH energy minimization seems promising, taking the best of the RH and DSM energy models, and improving the convergence compared to TRSCF, which already saw better or as good convergence rates as DIIS. It could be expected that this scheme has the ability to converge in fewest SCF iterations overall. The future success of ARH is dependent on the development of effective ways of solving the nonlinear equations in $\mathbf{X}$, e.g. by setting up a good preconditioner.

1.5 The Quality of the Energy Models for HF and DFT

Having considered the theory behind the TRRH and TRDSM steps in Section 1.4.1 and 1.4.2 without being concerned with the approximations introduced in the energy functions, this section takes a closer look at the errors in the energy models compared to the SCF energy. The SCF optimization of Hartree-Fock and Kohn-Sham-DFT energies is similar; the only difference lies in the energy expressions to be optimized. The approximations in the energy models will thus also differ in HF and DFT, and while Section 1.2 described the HF and DFT theory in a generic manner, this section will focus on the differences, ignoring the general elements already stated in Section 1.2.

To make the differences in the HF and DFT energy expressions clear, we will now study them separately:

$$E_{\mathrm{HF}} = 2\,\mathrm{Tr}\,\mathbf{hD} + \mathrm{Tr}\,\mathbf{DG}_{\mathrm{HF}}(\mathbf{D}) + h_{\mathrm{nuc}}\,, \tag{1.85}$$

$$E_{\mathrm{DFT}} = 2\,\mathrm{Tr}\,\mathbf{hD} + \mathrm{Tr}\,\mathbf{DG}_{\mathrm{DFT}}(\mathbf{D}) + h_{\mathrm{nuc}} + E_{\mathrm{XC}}(\mathbf{D})\,, \tag{1.86}$$

where

$$[\mathbf{G}_{\mathrm{HF}}(\mathbf{D})]_{\mu\nu} = 2\sum_{\rho\sigma} g_{\mu\nu\rho\sigma} D_{\rho\sigma} - \sum_{\rho\sigma} g_{\mu\sigma\rho\nu} D_{\rho\sigma}\,, \tag{1.87}$$

$$[\mathbf{G}_{\mathrm{DFT}}(\mathbf{D})]_{\mu\nu} = 2\sum_{\rho\sigma} g_{\mu\nu\rho\sigma} D_{\rho\sigma} - \gamma\sum_{\rho\sigma} g_{\mu\sigma\rho\nu} D_{\rho\sigma}\,. \tag{1.88}$$

The second term in Eq. (1.87) and Eq. (1.88) is the contribution from exact exchange, with $\gamma = 0$ in pure DFT (LDA), and $\gamma \neq 0$ in hybrid DFT. The exchange-correlation energy $E_{\mathrm{XC}}(\mathbf{D})$ in Eq. (1.86) is a functional of the electronic density. In the local-density approximation (LDA), the exchange-correlation energy is local in the density, whereas in the generalized gradient approximation (GGA), it is also local in the squared density gradient, and may thus be expressed as

$$E_{\mathrm{XC}}(\mathbf{D}) = \int f(\rho(\mathbf{x}), \zeta(\mathbf{x}))\,\mathbf{dx}\,. \tag{1.89}$$

Here the electron density $\rho(\mathbf{x})$ and its squared gradient norm $\zeta(\mathbf{x})$ are given by

$$\begin{aligned} \rho(\mathbf{x}) &= \boldsymbol{\chi}^{\mathrm{T}}(\mathbf{x})\mathbf{D}\boldsymbol{\chi}(\mathbf{x})\,, \\ \zeta(\mathbf{x}) &= \nabla\rho(\mathbf{x})\cdot\nabla\rho(\mathbf{x})\,, \end{aligned} \tag{1.90}$$

where $\boldsymbol{\chi}(\mathbf{x})$ is a column vector containing the AOs. Note that the exchange-correlation energy density $f(\rho(\mathbf{x}), \zeta(\mathbf{x}))$ in Eq. (1.89) is a nonlinear (and non-quadratic) function of $\rho(\mathbf{x})$ and $\zeta(\mathbf{x})$. In the following is relied on an expansion of $E_{\mathrm{XC}}(\mathbf{D})$ around some reference density matrix $\mathbf{D}_0$

$$E_{\mathrm{XC}}(\mathbf{D}) = E_{\mathrm{XC}}(\mathbf{D}_0) + (\mathbf{D}-\mathbf{D}_0)^{\mathrm{T}}\,\mathbf{E}_{\mathrm{XC}}^{(1)} + \tfrac{1}{2}(\mathbf{D}-\mathbf{D}_0)^{\mathrm{T}}\,\mathbf{E}_{\mathrm{XC}}^{(2)}(\mathbf{D}-\mathbf{D}_0) + \cdots\,, \tag{1.91}$$

where the derivatives $\mathbf{E}_{\mathrm{XC}}^{(n)}$ have been evaluated at $\mathbf{D} = \mathbf{D}_0$ and where for convenience a vector-matrix notation for $\mathbf{D}$, $\mathbf{E}_{\mathrm{XC}}^{(1)}$, and $\mathbf{E}_{\mathrm{XC}}^{(2)}$ is used. The precise form of E_{XC} depends on the DFT functional chosen for the calculation.

It is often more problematic to obtain convergence for DFT than HF, mainly for two reasons: The HOMO-LUMO gap $\Delta\varepsilon_{ai}$ is smaller for DFT than for HF, and a determinant with a well separated occupied and virtual part has better convergence properties than one with a lot of close lying states[54,55]. Also, since the exchange-correlation is nonlinear and non-quadratic in the density, the higher order terms in the density not present in Hartree-Fock theory introduces some extra approximations to the SCF scheme for DFT. In this section these differences and their consequences for the convergence properties will be discussed for the TRSCF algorithm. It is here assumed that if the energy models employed in TRSCF were of the same quality for HF and DFT, that is, had errors

of the same order compared to the true SCF energy, then the convergence properties would also be of the same quality.

The study is mainly performed in the MO basis with a block diagonal Fock matrix as in Eq. (1.10) and the reference density matrix $\mathbf{D}_0^{\mathrm{MO}}$

$$\mathbf{D}_0^{\mathrm{MO}} = \begin{pmatrix} 2\delta_{ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{1.92}$$

It is also exploited that any valid density matrix $\mathbf{D}$ may be expressed in terms of a valid reference density matrix $\mathbf{D}_0$ as

$$\mathbf{D}^{\mathrm{MO}}(\mathbf{K}) = \exp(-\mathbf{K})\mathbf{D}_0^{\mathrm{MO}}\exp(\mathbf{K}), \tag{1.93}$$

and can thus be expanded in orders of $\mathbf{K}$ through the BCH-expansion[46]

$$\mathbf{D}^{\mathrm{MO}}(\mathbf{K}) = \mathbf{D}_0^{\mathrm{MO}} + \left[\mathbf{D}_0^{\mathrm{MO}}, \mathbf{K}\right] + \tfrac{1}{2}\left[\left[\mathbf{D}_0^{\mathrm{MO}}, \mathbf{K}\right], \mathbf{K}\right] + \mathcal{O}(\mathbf{K}^3). \tag{1.94}$$

The anti-symmetric rotation matrix may be written in the form

$$\mathbf{K} = \begin{pmatrix} \mathbf{0} & -\boldsymbol{\kappa}^{\mathrm{T}} \\ \boldsymbol{\kappa} & \mathbf{0} \end{pmatrix}, \tag{1.95}$$

where $\boldsymbol{\kappa}$ holds the orbital rotation parameters. The diagonal block matrices representing rotations among the occupied MOs and among the virtual MOs are zero since the density matrix in Eq. (1.8) is invariant to such rotations.

In the following subsections the RH energy model Eq. (1.22) and the DSM energy model Eq. (1.55) are analyzed separately with respect to differences for HF and DFT.

1.5.1 The Quality of the TRRH Energy Model

To compare the RH energy model to the SCF energy, both are expanded about a reference density matrix $\mathbf{D}_0$ (neglecting the possible difference between $\mathbf{F}_0$ and $\mathbf{F}(\mathbf{D}_0)$ noted in Section 1.4)

$$E^{\mathrm{RH}}(\mathbf{D}) = E^{\mathrm{RH}}(\mathbf{D}_0) + 2\,\mathrm{Tr}\,\mathbf{F}(\mathbf{D}_0)(\mathbf{D}-\mathbf{D}_0), \tag{1.96}$$

$$\begin{aligned} E_{\mathrm{SCF}}(\mathbf{D}) = {} & E_{\mathrm{SCF}}(\mathbf{D}_0) + 2\,\mathrm{Tr}\,\mathbf{F}(\mathbf{D}_0)(\mathbf{D}-\mathbf{D}_0) + \mathrm{Tr}(\mathbf{D}-\mathbf{D}_0)\mathbf{G}(\mathbf{D}-\mathbf{D}_0) \\ & + E_{\mathrm{XC}}(\mathbf{D}) - E_{\mathrm{XC}}(\mathbf{D}_0) - \mathrm{Tr}(\mathbf{D}-\mathbf{D}_0)\mathbf{E}_{\mathrm{XC}}^{(1)}(\mathbf{D}_0), \end{aligned} \tag{1.97}$$

where the last three terms of Eq. (1.97) only are present in DFT theory. These expansions have the same first-order term $2\mathrm{Tr}\mathbf{F}(\mathbf{D}_0)(\mathbf{D} - \mathbf{D}_0)$ and thus the same first derivative with respect to the orbital rotation parameters κ_{ai} of Eq. (1.95)

$$\left[\mathbf{E}_{\mathrm{RH}}^{(1)}\right]_{ai} = \left.\frac{\partial E^{\mathrm{RH}}(\boldsymbol{\kappa})}{\partial \kappa_{ai}}\right|_{\boldsymbol{\kappa}=0} = -4F_{ai}, \tag{1.98}$$

$$\left[\mathbf{E}_{\mathrm{SCF}}^{(1)} \right]_{ai} = \left. \frac{\partial E_{\mathrm{SCF}}(\boldsymbol{\kappa})}{\partial \kappa_{ai}} \right|_{\boldsymbol{\kappa}=0} = -4F_{ai} \,. \tag{1.99}$$

The expressions are found replacing **D** in Eqs. (1.96) and (1.97) with $\mathbf{D}^{\mathrm{MO}}$ in Eq. (1.94) and differentiating with respect to κ_{ai}.

All higher order terms in $\boldsymbol{\kappa}$ arising from $2\mathrm{Tr}\mathbf{F}(\mathbf{D}_0)(\mathbf{D} - \mathbf{D}_0)$ are consequently also shared for the SCF and RH energies whereas terms of second and higher order arising from the last term(s) in Eq. 1.94 are neglected in the RH energy model. To study the differences, the second order derivatives in $\boldsymbol{\kappa}$ are found in the same way as the first derivatives

$$\left[\mathbf{E}_{\mathrm{RH}}^{(2)} \right]_{aibj} = \left. \frac{\partial^2 E^{\mathrm{RH}}(\boldsymbol{\kappa})}{\partial \kappa_{ai} \partial \kappa_{bj}} \right|_{\boldsymbol{\kappa}=0} = 4\delta_{ij}\delta_{ab}\left(\varepsilon_a - \varepsilon_i \right) \tag{1.100}$$

$$\left[\mathbf{E}_{\mathrm{SCF}}^{(2)} \right]_{aibj} = \left. \frac{\partial^2 E_{\mathrm{SCF}}(\boldsymbol{\kappa})}{\partial \kappa_{ai} \partial \kappa_{bj}} \right|_{\boldsymbol{\kappa}=0} = 4\delta_{ij}\delta_{ab}\left(\varepsilon_a - \varepsilon_i \right) + W_{aibj} \,, \tag{1.101}$$

where

$$W_{aibj}^{\mathrm{HF}} = 16g_{aibj} - 4\left(g_{abij} + g_{ajib} \right) \tag{1.102}$$

$$W_{aibj}^{\mathrm{DFT}} = 16g_{aibj} - 4\gamma\left(g_{abij} + g_{ajib} \right) + \left[\mathbf{E}_{\mathrm{XC}}^{(2)}(\boldsymbol{\kappa}) \right]_{aibj} . \tag{1.103}$$

$\mathbf{E}_{\mathrm{XC}}^{(2)}(\boldsymbol{\kappa})$ is the second derivative of the term E_{XC} expanded in the orbital rotation parameters $\boldsymbol{\kappa}$. The error in the RH energy model can then be said to depend partly on the size of **W** and partly on the size of the third and higher order contributions from the nonlinear terms in Eq. (1.97) which are not included in Eq. (1.96). This general consideration goes for DFT as well as HF, but with different impact. As seen in Eq. (1.102) and (1.103), the definition of **W** differs in the two approaches and even differs depending on which DFT functional is chosen. Furthermore, since the size of the HOMO-LUMO gap $\Delta\varepsilon_{ai} = \varepsilon_a - \varepsilon_i$ is typically smaller in DFT, the term $4\delta_{ij}\delta_{ab}(\varepsilon_a - \varepsilon_i)$ will have different weights in Eq. (1.101) depending on the method. Also the size of the third and higher order contributions in Eq. (1.97) would be expected to differ for HF and DFT, since for DFT both the terms $\mathrm{Tr}(\mathbf{D} - \mathbf{D}_0)\mathbf{G}(\mathbf{D} - \mathbf{D}_0)$ and $E_{\mathrm{XC}}(\mathbf{D})$ contribute whereas HF only contains the $\mathrm{Tr}(\mathbf{D} - \mathbf{D}_0)\mathbf{G}(\mathbf{D} - \mathbf{D}_0)$ term. In the beginning of the optimization, where large steps are taken, the size of the third and higher order contributions is the potential source of error. Near convergence this should be less of an issue, and in this region the size of the lowest Hessian eigenvalues should be the decisive error source.

HF and LDA calculations have been carried out and the part of the SCF energy change arising from the RH step $\Delta E_{\mathrm{SCF}}^{\mathrm{RH}}$ has been found as well as the change in the RH energy model ΔE^{RH} in each iteration.

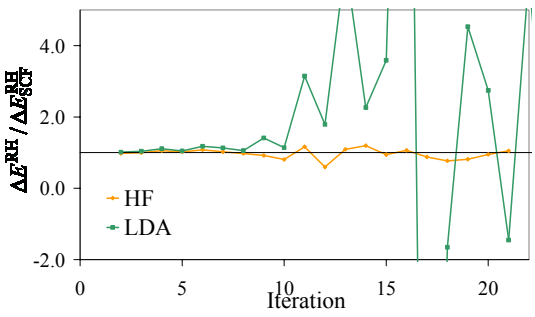


Fig. 1.27 Calculations on the cadmium complex in Fig. 1.6 in the STO-3G basis set.

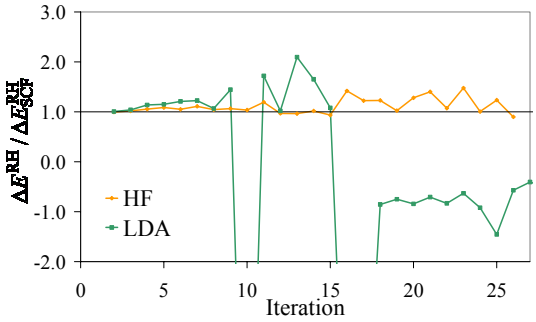


Fig. 1.28 Calculations on the zinc complex in Fig. 1.3 in the 6-31G basis set.

The change in the RH energy model is found as

$$\Delta E^{\mathrm{RH}} = 2\,\mathrm{Tr}\,\mathbf{F}\left(\mathbf{D}_{n+1} - \mathbf{D}_0^{\mathrm{idem}}\right), \tag{1.104}$$

where $\mathbf{D}_0^{\mathrm{idem}}$ is the reference density matrix, typically a $\bar{\mathbf{D}}$ from the previous TRDSM step purified as in Eqs. (1.32)-(1.33), and $\mathbf{D}_{n+1}$ is the new density found from diagonalization of the Fock matrix. In the **C**-shift scheme the criterion Eq. (1.31) ensures that the occupied and virtual orbitals do not mix, and thus the Hessian, Eq. (1.100), is positive and the RH energy decreases. The SCF energy change is found as

$$\Delta E_{\mathrm{SCF}}^{\mathrm{RH}} = E_{\mathrm{SCF}}(\mathbf{D}_{n+1}) - E_{\mathrm{SCF}}(\mathbf{D}_0^{\mathrm{idem}}). \tag{1.105}$$

The ratio between Eq. (1.104) and Eq. (1.105) contains information of the quality of the RH energy model. If the errors are negligible, the ratio is close to 1. If the ratio is larger than one, the RH energy model exaggerates the energy decrease, and if it is between 0 and 1 it underestimates the energy decrease. If it is negative, the SCF energy increases even though the RH energy model predicts an energy decrease.

For two test cases the $\Delta E^{\mathrm{RH}} / \Delta E_{\mathrm{SCF}}^{\mathrm{RH}}$ ratio is displayed in Fig. 1.27 and Fig. 1.28, respectively. It is clearly seen that generally, the RH energy model is better for HF than for DFT, in particular, negative values are seen for the LDA ratios. The errors in the RH energy model for the LDA calculations get worse as convergence is approached, so it would be expected that the significant source of error is the neglected term **W** in the Hessian rather than the higher order terms. Since locally the lowest Hessian eigenvalue should be the one controlling the optimization, this theory is inspected evaluating the lowest Hessian eigenvalue for both the RH energy model and for SCF according to Eq. (1.100) and Eq. (1.101), respectively, at convergence of the two test cases. The results are compared in Table 1-4.

Table 1-4 The lowest Hessian eigenvalues for the RH energy model and SCF energy at convergence of the calculations in Fig. 1.27 and Fig. 1.28. The deviation is found as $\left(\left[\mathbf{E}_{\mathrm{RH}}^{(2)}\right]_{\min} - \left[\mathbf{E}_{\mathrm{SCF}}^{(2)}\right]_{\min}\right)\cdot 100\% \Big/ \left[\mathbf{E}_{\mathrm{SCF}}^{(2)}\right]_{\min}$.

	cadmium complex		zinc complex	
	HF	LDA	HF	LDA
$\left[\mathbf{E}_{\mathrm{SCF}}^{(2)}\right]_{\min}$	0.557	0.017	1.000	0.290
$\left[\mathbf{E}_{\mathrm{RH}}^{(2)}\right]_{\min}$	1.112	0.014	1.621	0.281
Deviation	100%	-21%	62%	-2%

As expected, the lowest Hessian eigenvalue for the RH energy model, that is the HOMO-LUMO gap, is much smaller for LDA than for HF, but surprisingly it is seen that the Hessian prediction in the RH energy model for LDA is much better than the one for HF. Of course this is only the lowest eigenvalue, and we have not studied the corresponding eigenvector. We know for sure that the size of the orbital rotation parameters κ_{ai} decreases during the optimization and should be very small at convergence, where only small adjustments to the density are made. It is thus difficult to imagine that terms of third and higher order in $\boldsymbol{\kappa}$ should be the reason for the larger errors in the DSM energy model for LDA compared to HF.

This is a matter we will investigate further in the future since it is not understood at the moment. The importance of the higher order terms should be examined directly to understand how they affect the errors, and the Hessian should be studied more carefully introducing information about the direction of the eigenvalues. However, it can still be concluded from Fig. 1.27 and Fig. 1.28 that the RH energy model is poorer for LDA than for HF optimizations.

1.5.2 The Quality of the TRDSM Energy Model

The TRDSM energy model of Section 1.4.2.2 is formulated in a general manner and is as applicable to DFT theory as to HF theory. Still, the model will be poorer for DFT than for HF because of the general exchange-correlation term appearing in the DFT energy.

For the DSM energy model there are in general four possible sources of errors:

1. The purified density $\tilde{\mathbf{D}}$ still has an idempotency error.

2. The term $\frac{1}{2}\mathbf{D}_{\delta}^{\mathrm{T}}\mathbf{E}_{0}^{[2]}\mathbf{D}_{\delta}$ in $E(\tilde{\mathbf{D}})$, Eq. (1.50), is neglected.

3. $E(\tilde{\mathbf{D}})$, Eq. (1.50), is truncated after second order.

4. $\mathbf{E}_{0}^{(2)}\mathbf{D}_{+}$ in Eq. (1.50) is approximated by $2\mathbf{F}_{+}$.

Let us take a closer look at the errors one by one. In ref. [39] a general order analysis of the purified density $\tilde{\mathbf{D}}$ used in the parameterization of the DSM energy is given, and the results are summarized in Table 1-5.

Table 1-5. Comparison of the properties of the unpurified density $\bar{\mathbf{D}}$ and the purified density $\tilde{\mathbf{D}}$. c is the density expansion coefficients and $\boldsymbol{\kappa}$ is the orbital rotation parameters that change $\mathbf{D}_0$ to another density in the subspace $\mathbf{D}_i$.

	$\bar{\mathbf{D}}$	$\tilde{\mathbf{D}}$
Differences	$\mathbf{D}_+ = \bar{\mathbf{D}} - \mathbf{D}_0 = \mathcal{O}(c\|\boldsymbol{\kappa}\|)$	$\mathbf{D}_\delta = \tilde{\mathbf{D}} - \bar{\mathbf{D}} = \mathcal{O}\left(c\|\boldsymbol{\kappa}\|^2\right)$
Idempotency error	$\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}} - \bar{\mathbf{D}} = \mathcal{O}\left(c\|\boldsymbol{\kappa}\|^2\right)$	$\tilde{\mathbf{D}}\mathbf{S}\tilde{\mathbf{D}} - \tilde{\mathbf{D}} = \mathcal{O}\left(c^2\|\boldsymbol{\kappa}\|^4\right)$
Trace error	$\mathrm{Tr}\,\bar{\mathbf{D}}\mathbf{S} - N/2 = 0$	$\mathrm{Tr}\,\tilde{\mathbf{D}}\mathbf{S} - N/2 = \mathcal{O}\left(c^2\|\boldsymbol{\kappa}\|^4\right)$

In the $\tilde{\mathbf{D}}$ column, the order of the idempotency correction $\mathbf{D}_\delta$ and the idempotency error for $\tilde{\mathbf{D}}$ are found. These are the same for DFT and HF; the idempotency error is of order $c^2||\boldsymbol{\kappa}||^4$, and since $\mathbf{D}_\delta$ is of the order $c||\boldsymbol{\kappa}||^2$, the error connected to the neglect of the term second order in $\mathbf{D}_\delta$, will be of order $c^2||\boldsymbol{\kappa}||^4$ as well.

The third possible source of errors is the truncation of the energy $E(\tilde{\mathbf{D}})$ after second order in the density. Since the Hartree-Fock energy is quadratic in the density, this truncation leads to no errors for HF, but for DFT there will be an error of order $||\mathbf{D}_+||^3$ and from the first column in Table 1-5 it is seen that it can be written as an error of order $c^3||\boldsymbol{\kappa}||^3$, since $\mathbf{D}_+$ is of the order $c||\boldsymbol{\kappa}||$. Also since the HF energy is quadratic in the density, no third derivative $\mathbf{E}_0^{(3)}$ exists and thus the Taylor expansion used to find $\mathbf{E}_0^{(2)}\mathbf{D}_+ = 2\mathbf{F}_+$ is terminated for HF, but for DFT terms of order $||\mathbf{D}_+||^2$ are neglected. Since $\mathbf{E}_0^{(2)}\mathbf{D}_+$ is multiplied by $\mathbf{D}_+$ in the energy function Eq. (1.50), this gives an error for DFT of the order $||\mathbf{D}_+||^3$ or as before $c^3||\boldsymbol{\kappa}||^3$. The sizes of the introduced errors are summarized in Table 1-6.

Table 1-6. Comparison of the errors introduced in the DSM energy model for HF and DFT respectively.

		error in HF	error in DFT
1	Idempotency error $\tilde{\mathbf{D}}\mathbf{S}\tilde{\mathbf{D}} - \tilde{\mathbf{D}}$	$\mathcal{O}\left(c^2\|\boldsymbol{\kappa}\|^4\right)$	$\mathcal{O}\left(c^2\|\boldsymbol{\kappa}\|^4\right)$
2	Neglected term $\frac{1}{2}\mathbf{D}_\delta^{\mathrm{T}}\mathbf{E}_0^{[2]}\mathbf{D}_\delta$	$\mathcal{O}\left(c^2\|\boldsymbol{\kappa}\|^4\right)$	$\mathcal{O}\left(c^2\|\boldsymbol{\kappa}\|^4\right)$
3	Truncation of $E(\tilde{\mathbf{D}})$	0	$\mathcal{O}\left(c^3\|\boldsymbol{\kappa}\|^3\right)$
4	Approximation of $\mathbf{E}_0^{(2)}\mathbf{D}_+$	0	$\mathcal{O}\left(c^3\|\boldsymbol{\kappa}\|^3\right)$

Depending on the sizes of c and $||\boldsymbol{\kappa}||$ respectively, the error for DFT will be of same or lower order than the one for HF. To inspect whether or not the DSM energy is a poorer model for DFT than for HF, a number of calculations have been carried out, and the sizes of $||\mathbf{D}_\delta||$ and $||\mathbf{D}_+||$ for the DSM step in each iteration are examined. Since $\mathbf{D}_\delta$ is of the order $c||\boldsymbol{\kappa}||^2$ and $\mathbf{D}_+$ is of the order $c||\boldsymbol{\kappa}||$, the

size of $||\mathbf{D}_\delta||^2$ and $||\mathbf{D}_+||^3$ will indicate whether the error in the energy model is controlled by the $\mathcal{O}\left(c^2\|\boldsymbol{\kappa}\|^4\right)$ or the $\mathcal{O}\left(c^3\|\boldsymbol{\kappa}\|^3\right)$ error. The test cases showed similar behavior and results from HF and LDA calculations on the cadmium complex in Fig. 1.6 with a STO-3G basis and a H1-core start guess are displayed in Fig. 1.29 and Fig. 1.30.

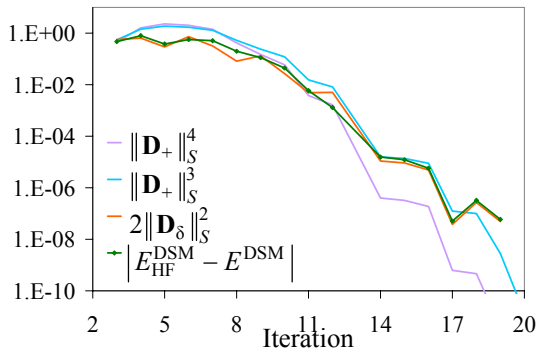


Fig. 1.29 HF/STO-3G calculation. The size of different density norms compared to the actual error in the DSM energy model.

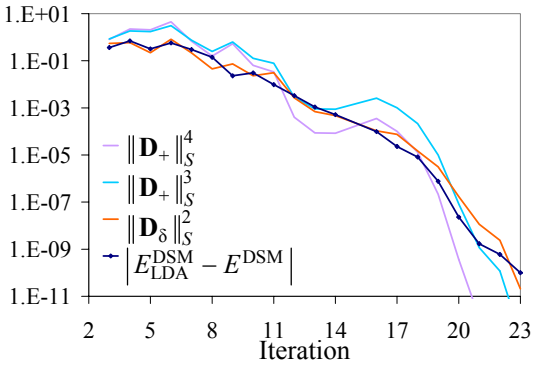


Fig. 1.30 LDA/STO-3G calculation. The size of different density norms compared to the actual error in the DSM energy model.

The SCF energy at the end of a DSM step $E_{\mathrm{SCF}}^{\mathrm{DSM}}$ is found by purifying the resulting $\bar{\mathbf{D}}$ by Eq. (1.32) –(1.33) and evaluating the SCF energy, Eq. (1.1), for this density. The DSM energy, Eq. (1.55), is also evaluated and the error of the DSM energy model is then found as the size $\left|E_{\mathrm{SCF}}^{\mathrm{DSM}} - E^{\mathrm{DSM}}\right|$. For the HF calculation this error is expected to be of the size $||\mathbf{D}_\delta||^2$, and it is seen in Fig. 1.29 that this is actually the case; if $||\mathbf{D}_\delta||^2$ is multiplied by 2, there is a remarkable fit. Also it is seen that if the error in the DSM energy for HF should be expressed in the density differences $\mathbf{D}_+$, it would be the density differences to the third rather than the fourth order. For the DFT calculation the interesting point was to see whether or not $||\mathbf{D}_+||^3$ is the controlling error. In Fig. 1.30 is seen that even though there is not an obvious fit as for HF, $||\mathbf{D}_\delta||^2$ seems to be the dominant error here as well. Still, if the error should be expressed in the density differences $\mathbf{D}_+$, it would be the density differences to the third rather than the fourth order as expected for DFT.

In conclusion it seems that the dominating error in the DSM energy both for HF and DFT is $||\mathbf{D}_\delta||^2$, that is, the idempotency correction squared. In comparison it should be mentioned that the EDIIS model[37] by Kudin, Scuseria, and Cancès corresponds to $E(\bar{\mathbf{D}})$ in Eq. (1.55) and thus has an error of the order $||\mathbf{D}_\delta||$ compared to the SCF energy.

1.6 Convergence for Problems with Several Stationary Points

The HF equation is a nonlinear equation and, therefore, it presents in principle several solutions. Several minima might exist, and even though it is typically preferred to find the global minimum,

no optimization method can make that a guarantee. Furthermore, it cannot be tested if the minimum found is a local or the global minimum without knowledge of the whole surface. Depending on the start guess and the optimization approach, an optimization can converge to different stationary points. Further, it is necessary to decide in which subspace of orbital rotations the desired solution should be found, since a solution representing a stable stationary point in one subspace is not necessarily stable in another.

Orbital rotations can be divided in real and complex rotations and each of those can be further divided in singlet and triplet rotations. Each of those can then again be divided in rotations within the different point group symmetries. Generally, we do not consider the complex rotations, and we only optimize in the real space. Further, when optimizing a closed shell wave function, only the total-symmetric part of the singlet rotations is considered. A stationary point in the subspace of real, total-symmetric, singlet rotations can be shown through elementary arguments to be a stationary point for all types of rotations. However, a stationary point can both be a maximum, a saddle point or a minimum. A way to realize if the stationary point also is a minimum is to evaluate the Hessian eigenvalues. This is done within the subspace in which the solution should be stable. If a negative Hessian eigenvalue is found in the subspace of singlet rotations, the stationary point is said to have a singlet instability and if a negative Hessian eigenvalue is found in the subspace of triplet rotations, it is said to have a triplet instability[54,56]. Triplet instabilities are connected to breaking the symmetry between α and β orbitals. If a triplet instability is found, a minimum with a lower energy than the current stationary point can be found, if the α and β parts are allowed to differ, typically leading to a solution which is not an eigenfunction of $\hat{S}^2$. Hence, the lower minimum could be found by an unrestricted HF (UHF) optimization. A singlet instability found in the total-symmetric subspace indicates that the current stationary point is a saddle point and a minimum with lower energy exists within the subspace. If a singlet instability is found outside the total-symmetric subspace, orbitals of different symmetries should be mixed to decrease the energy further, changing the symmetry of the resulting wave function.

The *aufbau* ordering rule assumes that occupying the orbitals of lowest energy also leads to the lowest Hartree-Fock energy. This cannot be proven to always apply for restricted HF as it can for UHF[57]. Thus it is a risk when the *aufbau* ordering is forced upon an optimization, that a lower energy with the *aufbau* ordering broken could exist. However in a study by Dardenne et. al.[58], in which different ordering schemes were tested, they found in all cases that the minimum was an *aufbau* solution. The *aufbau* ordering was broken only for saddle points. In our schemes we always apply the *aufbau* ordering rule, but if the RH step is level shifted to the end of the optimization, it can force the convergence to a non-*aufbau* solution.

1.6.1 Walking Away from Unstable Stationary Points

As concluded in the previous section, the Hessian eigenvalues should be tested to make sure the optimized state is stable. This is expensive, so it is only done when it is expected that the problem has several stationary points. Depending on the desired solution, only the relevant part of the Hessian is checked. So far we have only considered singlet instabilities, but currently tests for triplet instabilities are implemented as well.

The check for singlet instabilities is made on the converged wave function, finding the lowest Hessian eigenvalue of the Hessian in the real, singlet subspace. If the lowest Hessian eigenvalue turns out to be positive, we are sure to have a solution which is stable with respect to singlet rotations, but if it is negative we are in a saddle point, and a minimum with a lower energy exists within the subspace. We have in our SCF program implemented the possibility to test the singlet Hessian and in case of a negative lowest Hessian eigenvalue follow the corresponding direction downhill and away from the saddle point. The scheme and some examples of its use will be described in the following.

1.6.1.1 Theory

When the SCF optimization has converged, the set of optimized orbitals described by their expansion coefficients $\mathbf{C}_{\text{opt}}$ are used to evaluate the lowest Hessian eigenvalues and the corresponding eigenvectors by an iterative subspace method. If the lowest Hessian eigenvalue ε_{min} is found positive, then it is clear that the optimization has converged to a minimum. If on the other hand the eigenvalue is negative, we know for sure that a lower stationary point exists.

We would then like to take a step downhill in the direction $\mathbf{x}$ corresponding to the negative eigenvalue ε_{min}

$$\mathbf{E}_{\text{SCF}}^{(2)}\mathbf{x} = \varepsilon_{\text{min}}\mathbf{x}\,. \tag{1.106}$$

This can be accomplished making a unitary transformation of the optimized expansion coefficients $\mathbf{C}_{\text{opt}}$ with $\mathbf{x}$ as the orbital rotation parameters to define the direction $\mathbf{X}_{\text{dir}}$ of the step

$$\mathbf{X}_{\text{dir}} = \begin{bmatrix} \mathbf{0} & -\mathbf{x}_{ai}^{\text{T}} \\ \mathbf{x}_{ai} & \mathbf{0} \end{bmatrix}. \tag{1.107}$$

The step length is controlled by a parameter α

$$\mathbf{U}_{\alpha} = \exp(-\alpha\mathbf{X}_{\text{dir}}) \tag{1.108}$$

$$\mathbf{C}'_{\text{opt}}(\alpha) = \mathbf{C}_{\text{opt}}\mathbf{U}_{\alpha}\,. \tag{1.109}$$

A line search is then carried out for $\alpha > 0$ to find the lowest SCF energy in the direction $\mathbf{X}_{\text{dir}}$. This is of course expensive since every point in the line search requires an evaluation of the Fock matrix

with respect to the new coefficients $\mathbf{C}'_{\text{opt}}$. When the SCF energy minimum in the direction $\mathbf{X}_{\text{dir}}$ is found, the corresponding coefficients should be the initial orbitals for a new SCF optimization, hopefully now optimizing further downhill to a minimum. In problematic cases, e.g. with a very flat saddle point close to the minimum, we have found it convenient to continue the optimization with the line search scheme TRSCF-LS (the combination of TRRH-LS and TRDSM-LS described in Sections 1.4.1.4 and 1.4.2.4) to ensure a continued decrease in the energy.

1.6.1.2 Examples

In Fig. 1.31 and Fig. 1.32 two examples of problems with several stationary points are given.

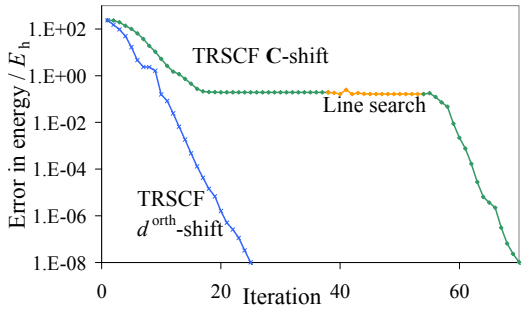


Fig. 1.31 HF calculations on the rhodium complex.

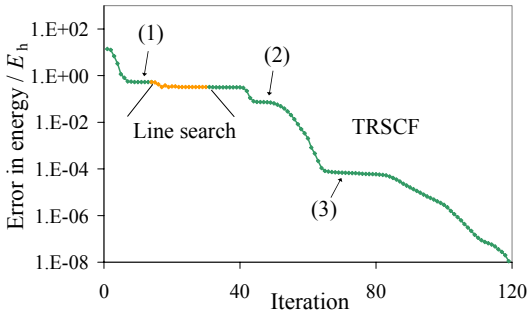


Fig. 1.32 HF/STO-3G calculation on CrC.

The first example is a HF optimization on the rhodium complex seen in Fig. 1.33 in the AhlrichsVDZ basis[59] combined with STO-3G on rhodium. For this example DIIS diverges, but the TRSCF scheme with **C**-shift converges nicely in 38 iterations. However, when the Hessian is inspected it is found that the lowest eigenvalue is negative, and a search in α is carried out in the direction corresponding to the negative eigenvalue. This is illustrated with the orange line in the picture. Since each evaluation of a step-length α necessitates an evaluation of the Fock matrix, it is fair to display each line search step as an iteration on the SCF iteration scale. When a minimum is found in this direction, the corresponding orbitals are used as a start guess for a new TRSCF optimization, and it is seen that it now converges nicely to a new and lower stationary point which is found to be a minimum. When the d^{orth}-shift scheme is applied in the TRRH steps instead of the **C**-shift scheme, it turns out that convergence to the minimum is obtained with no problems, as seen from Fig. 1.31, illustrating how the stationary point found from an SCF optimization not only depends on the start guess, but also on the optimization procedure.

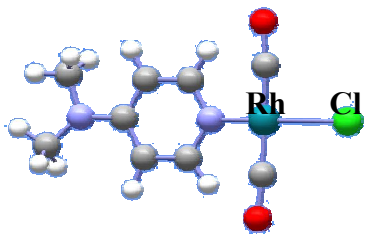


Fig. 1.33 Rhodium complex.

The second example is a HF/STO-3G optimization of CrC with a bond distance on 2.00Å. The example is also used in Fig. 1.13 and Fig. 1.25, but without discussing the stability of the converged state. Also in this case DIIS diverges whereas TRSCF converges nicely in 12-13 iterations to a stationary point which is found to have singlet instabilities. As for the first example, a line search is carried out in the downhill direction and a new TRSCF optimization is started from the resulting orbitals. This time the second optimization has more problems than was the case for the rhodium example, but finally it converges to a minimum. Whereas in the rhodium case, only one plateau corresponding to the saddle point could be seen, in this case three plateaus can be found, marked by numbers on the figure. The first is the saddle point that TRSCF converges to, at $E_{SCF} = -1068.77014939$ and with a lowest Hessian eigenvalue of -0.624. The second and third stationary points are recognized as saddle points by TRSCF itself and it manages to move away. If a DIIS optimization is carried out with a Hückel start guess, it converges to the second stationary point, which has E_{SCF} = -1069.21761813 and a lowest Hessian eigenvalue of -0.038, again demonstrating that depending on the optimization procedure and start guess, different stationary points can be found. It is thus necessary to check the Hessian of the result to know for sure that a minimum is found, and in this case the final minimum has E_{SCF} = -1069.30090709 and a lowest Hessian eigenvalue of 0.043. CrC is well known for being a molecule with a complicated electronic energy surface and has been the object for several theoretical studies[60].

The scheme testing for singlet instabilities and walking away from unstable stationary points could be integrated more efficiently in the optimization than is done here. It can be seen from Fig. 1.31 and Fig. 1.32 that the optimizations are completely converged before the Hessian check is made, spending many iterations improving the unwanted result. The check could be made in an earlier stage, saving a number of iterations. Also the steps taken in the line search could be optimized such that fewer steps were necessary to find the minimum. Anyhow, it is convenient to have the possibility to continue an optimization until a minimum is found.

1.7 Scaling

As mentioned in the introduction, it is now possible to apply *ab-initio* quantum chemical methods, in particular HF and DFT, to large molecular systems of interest for biology and nano-science. This is due to both the developments in integral screening and algorithms for the Fock matrix builder and to approaches avoiding diagonalization and exploiting sparsity in the matrices. Since the TRSCF scheme has properties which would be of great advantage for SCF calculations on large and complex molecules, it is crucial that the scheme can be formulated in a linear or near-linear scaling manner. We have not been concerned with the build of the Fock matrix, and any state-of-the-art, linear or near-linear scaling approach could be used as the Fock builder for our scheme. The steps to

consider are thus the Roothaan-Hall step TRRH, which evaluates a new density matrix, and the density subspace minimization TRDSM, which improves convergence. In the following subsections the scaling of these steps will be discussed.

1.7.1 Scaling of TRRH

The TRRH scheme with **C**-shift described in Section 1.4.1.2 requires the diagonalization of a level shifted Fock matrix and the knowledge of the occupied molecular orbital coefficients. The diagonalization scales as well as a matrix multiplication as N^3, where N is the dimension of the problem, in this case the number of basis functions. However, a diagonalization is ineffective and cannot be nearly as well optimized as a matrix multiplication, and thus the scaling factor is much larger for the diagonalization than for the matrix multiplication. Also, the matrix multiplication can exploit sparsity and obtain a scaling linearly in the number of non-zero elements whereas sparsity is not as easily exploited in diagonalizations. Furthermore, the molecular orbitals described by the eigenvectors from the diagonalization of the Fock matrix are inherently delocalized and thus there is no sparsity to exploit.

To obtain a linear scaling TRRH step it is thus necessary to avoid completely the diagonalizations and any reference to the MO basis. This can be done in our SCF program – a local version of DALTON[38,49] - by combining the d^{orth}-shift scheme described in Section 1.4.1.5 with the trace purification (TP) described in Section 1.4.1.6.

The trace purification scheme replaces the diagonalization of the level shifted Fock matrix and makes it possible to exploit sparsity in the matrices. A sparse blocked matrix storage scheme has been implemented for this purpose. In this scheme the columns and rows in the matrices are permuted such that close lying atoms are collected in blocks, making it possible to exploit the locality in the basis functions. Based on some drop tolerance for the size of matrix elements, pure zero blocks can be found and neglected, both saving storage and computing time. A library has been developed for the purpose of handling the matrix operations for this type of matrices and controlling the truncation error arising from the neglect of elements[49].

Calculations have been carried out on glycine chains of different length in the 4-31G basis set on a 3.4GHz Xeon/Nocona Machine with EM64T architecture and MKL BLAS+LAPACK library. Timings have been made in the third iteration of the SCF optimization, measuring how much time (CPU) is spent in the TRRH step in the case of full matrices and diagonalizations of the level shifted Fock matrix (Diag./full) and in the case of sparse blocked matrices and the TP scheme (TP/sparse). The results are seen in Fig. 1.34. Both in the full and sparse case the d^{orth}-shift scheme is applied.

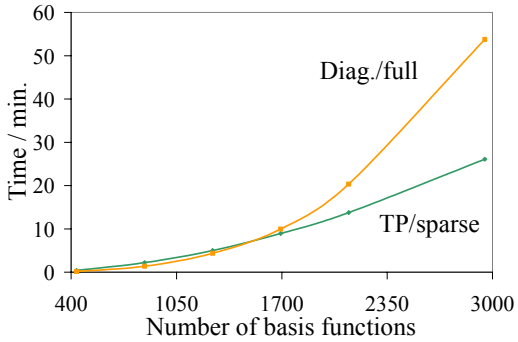


Fig. 1.34 Timings of a TRRH step in case of diagonalizations of full matrices (Diag./full) and in case of trace purification of sparse blocked matrices (TP/sparse).

The crossover is already around 1500 basis functions, and it is clear how the diagonalization scheme quickly will become too time consuming if the number of basis functions is increased further. Of course, this is a linear molecule as seen from Fig. 1.35, and the cross over will be later for more three-dimensional molecules. The TP method does not have an exact linear scaling because of the transformation to the orthogonal basis which gives rise to a quadratic term, but the scaling factor on the quadratic term is very small. It should be noted that the dynamic level shift scheme typically takes 5-10 diagonalizations or trace purifications to find the optimal level shift in the first couple of iterations, and as the timings are from the third iteration, then not just one, but several diagonalizations or purifications are included in the timings in Fig. 1.34. Currently a full trace purification optimization (30-70 purification iterations) is carried out for each level shift tested to find the optimal level shift. It is straightforward to optimize this process such that the purification is not converged as hard for the level shifts tested and rejected, as for the final optimal level shift.

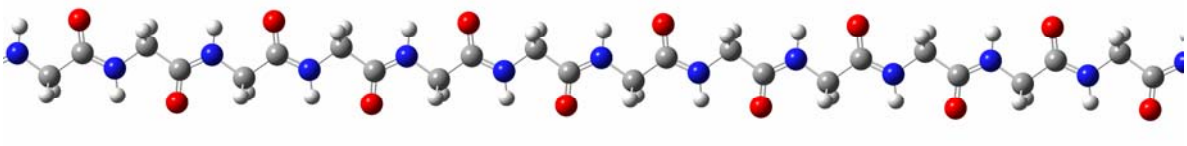


Fig. 1.35 Glycine chain.

To conclude, the scaling of the TRRH scheme with **C**-shift is dominated by the diagonalization, and sparsity cannot be exploited. Still with a good Fock builder it can run effectively up to a couple of thousand basis functions, but at some point the diagonalizations get too time consuming. For larger systems the purification scheme with the d^{orth}-shift scheme can be used with blocked sparse matrices resulting in a near-linear scaling.

1.7.2 Scaling of TRDSM

For the density subspace minimization, a set of linear equations, Eq. (1.66), are solved in each DSM step, but only in the dimension of the subspace which is much smaller than the number of basis functions. It is therefore of no significance compared to the matrix additions and multiplications needed to set up the DSM gradient **g** and Hessian **H** for the linear equations. For TRDSM it will thus only be the number of matrix multiplication that determines the scaling. Nothing has to be changed to exploit sparsity in the matrices, and linear scaling is automatically obtained from the point where the number of non-zero elements in the matrices is linear scaling. For full matrices the scaling is formally N^3, where N is the number of basis functions, but as mentioned in the previous subsection this is not a problem as it is for the diagonalization, since matrix multiplications can be carried out with close to peak performance on computers. However, the number of matrix multiplications should be kept at a minimum as it affects the scaling factor.

The number of matrix multiplications is dependent on the dimension of the subspace as the number of gradient and Hessian elements grows with the size of the subspace, but even though the Hessian is set up explicitly, the number of matrix multiplications only scales linearly with the dimension of the subspace. The expressions for the DSM gradient and Hessian are found in 0, and it is seen that if only the matrices $\bar{\mathbf{F}}\mathbf{D}_i$, $\mathbf{S}\mathbf{D}_i$, $\bar{\mathbf{F}}\mathbf{D}_i\mathbf{S}$ and $\bar{\mathbf{D}}\mathbf{S}\mathbf{D}_i$ are evaluated, then all the terms for a Hessian element can be expressed as the trace of two known matrices or their transpose. As the operation Tr**AB** scales quadratically instead of cubically, the overall scaling of TRDSM will be nN^3 for full matrices, where n is the dimension of the subspace and N the dimension of the problem. For sparse matrices both the matrix multiplications and Tr**AB** scale linearly, but since n^2 Tr**AB**s are evaluated, the overall scaling is n^2N. However, the trace operations have a very small prefactor.

In the TRSCF scheme with **C**-shift the diagonalizations are thus the dominating operations, but since both the TRRH and TRDSM step can be carried out without any reference to the MO basis and with matrix multiplications as the most expensive operations, the TRSCF scheme is near-linear scaling and has what it takes to be applied to really large molecular systems. It is still a work in progress to get all the parts working together, so unfortunately no large scale TRSCF calculations will appear in this thesis, and no benchmarks in which sparsity in the matrices is exploited for TRDSM can be presented, but the whole framework is in place.

1.8 Applications

In this section, numerical examples are given to illustrate the convergence characteristics of the TRSCF and ARH calculations. Comparisons are made with DIIS, the TRSCF-LS method, and the globally convergent trust-region minimization method (GTR) of Francisco *et. al.*[26].

In Section 1.8.1 a set of small molecules used by Francisco *et. al.* to illustrate the convergence characteristics of GTR is considered. Next in Section 1.8.2 the convergence of calculations on three metal complexes is discussed for the DIIS, TRSCF and TRSCF-LS methods.

1.8.1 Calculations on Small Molecules

As an alternative to the RH diagonalization, Francisco *et. al.* have developed an energy minimization method (GTR), where an energy model is minimized by a trust-region minimization. They have proven that it is a globally convergent algorithm, that is, no matter the starting point; the iterative steps will converge towards a stationary point. The best results are obtained when they combine GTR with DIIS and thereby let DIIS accelerate the convergence. To examine the convergence characteristics of TRSCF and ARH compared to GTR, calculations have been carried out with the attempt to reproduce the conditions given in the paper by Francisco *et. al.*. Thus HF calculations have been carried out with a maximum number of 10 previous density matrices for the density subspace minimizations and convergence is obtained when the difference between two consecutive energies is smaller than $10^{-9}E_h$. The results are given in Table 1-7; the numbers found with our SCF program are on a white background, whereas results copied from the GTR paper are on a grey background.

Table 1-7 Number of iterations in HF calculations performed by each algorithm in some test problems. The geometry of the molecules and the results in grey are taken from the paper by Francisco *et. al.*[26], and GTR+DIIS is their globally convergent trust-region algorithm with DIIS acceleration.

Molecule	Basis	Start guess	Algorithm					
			DIIS	TRSCF **C**-shift	TRSCF d^{orth}-shift	ARH	DIIS	GTR +DIIS
H_2O	STO-3G	H1-core	7	7	7	6	5	5
	6-31G	H1-core	10	9	8	8	8	8
NH_3	STO-3G	H1-core	7	8	7	6	7	7
	6-31G	H1-core	9	9	8	8	7	7
CO	STO-3G	H1-core	12	9	9	9	11	10
		Hückel	8	8	8	-	7	7
CO(Dist)*	STO-3G	H1-core	39(a)	9	8	8	117(b)	10
		Hückel	35	10	8	-	85	15
	6-31G	H1-core	24(a)	13	10	9	27(b)	115
		Hückel	21(a)	10	10	-	36(b)	59
Cr_2	STO-3G	H1-core	34(a)	14(a)	10(a)	12(a)	13	38
CrC	STO-3G	H1-core	29(a)	13(a)	11(a)	10(a)	(X)	29

* Distorted geometry – double bond length compared to CO
(a) Negative Hessian eigenvalue.
(b) Converged to a higher energy than some of the other algorithms
(X) No convergence in 5001 iterations.

Let us first consider the results obtained from our SCF program. Comparing the TRSCF results (both **C**-shift and d^{orth}-shift) to the DIIS results, it is clear that the TRSCF method not only is an

improvement when DIIS cannot converge, but also for small simple examples, the convergence of TRSCF is as good as or better than for DIIS. Also it is observed that in five instances DIIS converge to a stationary point which is not a minimum, while that only happens in two instances for TRSCF. This suggests that the TRSCF algorithm does not have a high tendency to converge to saddle points compared to DIIS. Comparing the results obtained for TRSCF with the **C**-shift and the d^{orth}-shift schemes, only minor differences are seen for these small examples, but in all cases the d^{orth}-shift scheme presents a faster or similar convergence rate compared to the **C**-shift scheme. With the ARH method the convergence is further improved compared to the TRSCF/d^{orth}-shift scheme. It is only a matter of saving a single iteration in some of the examples, but the tendency is clear. As the algorithm is still in the implementation phase, no numbers can currently be obtained with the Hückel start guess.

Comparing now the results from our SCF program with the results from the GTR paper, the obvious peculiarity is the discrepancies between the DIIS results obtained by Francisco *et. al.* and by us. A plain DIIS optimization should be completely reproducible, but there is a difference of two out of seven iterations. These differences cannot be explained and make it more difficult to compare our results with theirs. Furthermore it seems that they have not tested the Hessian eigenvalues at the end; only if they for some other start guess or optimization method found a lower energy, it is noted in their table, and thus we cannot know for sure if the given number of iterations corresponds to convergence to a minimum. For Cr_2 and CrC it is very difficult to find the minimum, and several saddle points exist where convergence can be obtained (see Section 1.6). It is thus an open question whether the GTR+DIIS calculations for Cr_2 and CrC actually converge to a minimum or to a saddle point as for the TRSCF methods.

In the examples where GTR+DIIS gives an improvement compared to their DIIS results, TRSCF and ARH also give significant improvements to our DIIS results. For the distorted CO example, TRSCF and ARH show better convergence than GTR+DIIS even if the results could be compared directly. For all examples TRSCF and ARH converge in 7-14 iterations, whereas GTR+DIIS use between five and 115. However, as discussed in Section 1.4.1.3, DIIS does not perform well when the gradient and energy are not correlated as is often the case in the global region when using TRRH, and could very well be the case for GTR as well. TRRH should be combined with a density subspace minimization method in the energy (e.g. TRDSM), and the same probably applies for GTR. We would thus suggest an implementation of TRDSM in connection with GTR.

In conclusion it has been illustrated that the TRSCF and ARH methods have very nice convergence properties with improvements compared to DIIS in general and to GTR+DIIS as well, in case of more problematic examples.

1.8.2 Calculations on Metal Complexes

In reference 39 and throughout this part of the thesis, three molecules including transition metals have been used for examples, namely the molecules in Fig. 1.3, Fig. 1.6 and Fig. 1.33. In this section HF and LDA calculations on these metal complexes are given both for DIIS, TRSCF and TRSCF-LS. For all calculations a H1-core start guess has been employed and a maximum of 10 matrices are used to define the subspace in the density subspace minimization. This is different from the examples given in ref. 39, where the subspace dimension never was larger than eight. Furthermore for the TRSCF calculations in ref. [39] the **C**-shift scheme was applied whereas in the calculations reported here, the d^{orth}-scheme has been applied.

TRSCF-LS is the TRSCF line search method in which the TRRH-LS and TRDSM-LS steps described in Sections 1.4.1.4 and 1.4.2.4 are combined to set up an expensive, but highly robust method, in which the lowest SCF energy is identified by a line search at each step. The convergence results of the optimizations are seen in Fig. 1.36. For the cadmium complex a STO-3G basis set has been applied, for the rhodium complex the AhlrichsVDZ basis set[59] has been applied except for the rhodium which is described in the STO-3G basis and for the zinc complex the 6-31G basis set has been applied.

The convergence of the TRSCF and TRSCF-LS methods is comparable for all cases in Fig. 1.36, and in general the TRSCF calculations converge in fewer iterations than the TRSCF-LS calculations do. As mentioned the line search method TRSCF-LS is much more expensive than TRSCF, and the only reason for applying it instead of TRSCF is for very difficult examples, where convergence cannot be obtained in any other way.

The convergence behavior of the DIIS method is somewhat more erratic than that of the TRSCF methods since it makes no use of Hessian information and therefore cannot predict reliably what directions will reduce the total energy. The HF calculation on the rhodium complex and the LDA calculation on the zinc complex both diverge for the DIIS method. In general the erratic behavior is in particular seen in the global region whereas in the local region, it converges as well as the TRSCF method.

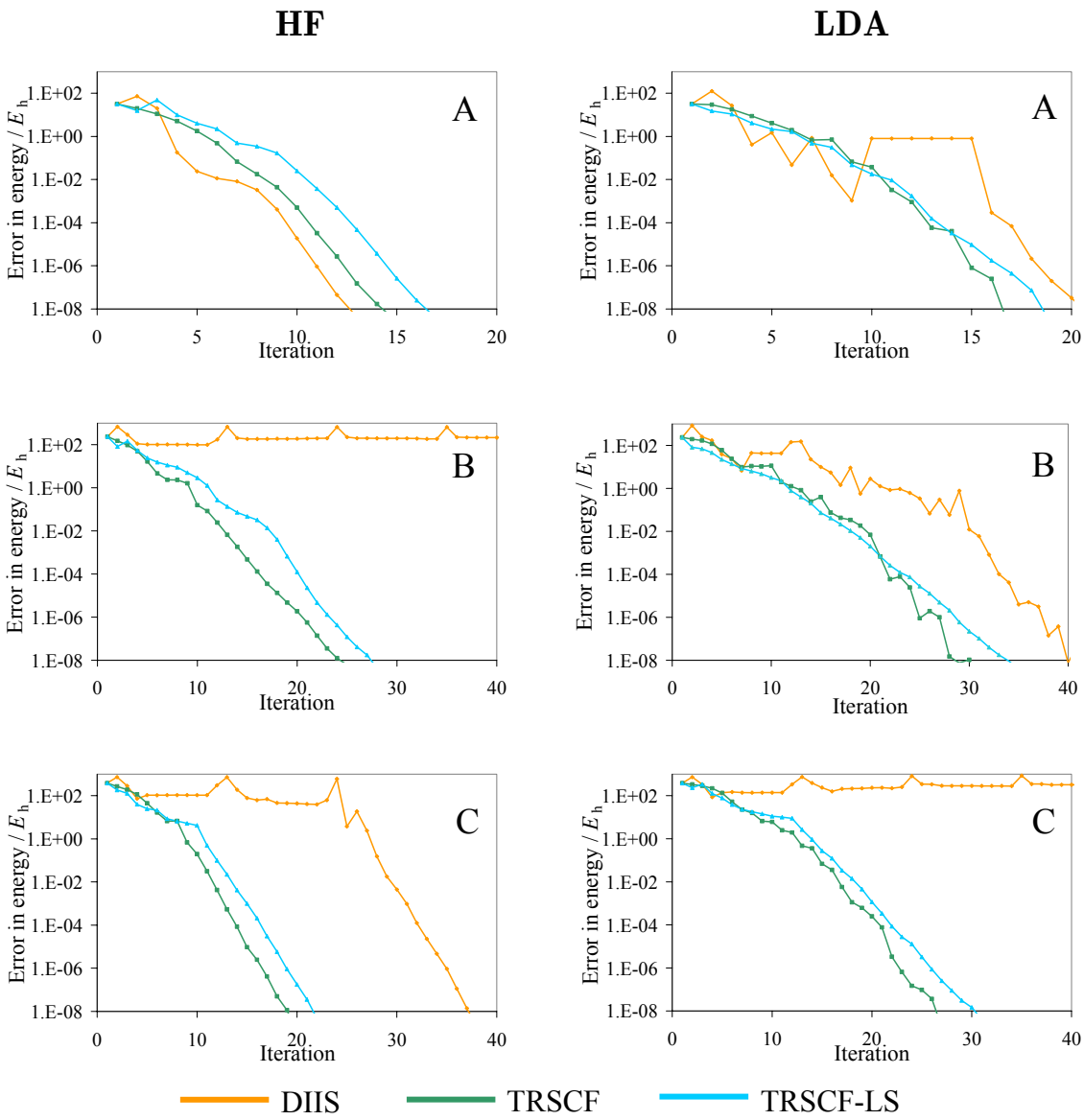


Fig. 1.36 Convergence of HF and LDA calculations on (A) the cadmium complex from Fig. 1.6, (B) the rhodium complex from Fig. 1.33, and (C) the zinc complex from Fig. 1.3.

For the examples presented both in this and the previous subsection, the TRSCF convergence is as good as or better than DIIS, and for problems where DIIS diverges, convergence is obtained with the TRSCF methods. It thus seems that TRSCF has the properties of a good black-box optimization algorithm.

1.9 Conclusion

In this part of the thesis the trust region SCF (TRSCF) algorithm is presented as a means to improve SCF convergence compared to methods typically used today e.g. DIIS. In the TRSCF method, both the Roothaan-Hall (RH) step and the density-subspace minimization (DSM) steps are replaced by optimizations of local energy models of the Hartree-Fock/Kohn-Sham energy E_{SCF}. These local models have the same gradient as the energy E_{SCF}, but an approximate Hessian. Restricting the steps of the TRSCF algorithm to the trust region of these local models, that is, to the region where the local models approximate E_{SCF} well, smooth and fast convergence may be obtained.

The developments through the years in SCF optimization algorithms are reviewed, and it is found that the fundamental schemes used in TRSCF to improve convergence have been around for several years; DIIS is actually a subspace minimization in the gradient norm, and level shifts have been used to improve or force convergence since 1973. Anyhow, the level shifts have previously been found on a trial and error basis as a constant parameter, whereas we advocate a dynamic level shift scheme in which the level shift is used to control the density change in the RH step. As such the level shift is optimized in each iteration to allow the density to change to the trust radius of the RH energy model, hence the name trust region Roothaan-Hall (TRRH) for our RH scheme. Also, the density subspace minimization has been improved compared to previous methods. An accurate energy model is constructed in the iterative subspace, where only minor approximations are made compared to the SCF energy. The trust region minimization of this energy model thus corresponds well to a minimization of E_{SCF} in the iterative subspace, thus resulting in an energy decrease in each trust region DSM (TRDSM) step. The TRRH and TRDSM steps in combination make up a successful scheme with a high convergence rate without compromising the control of the density changes in each step.

Compared to ref. [38] and [39], an alternative level shift scheme (d^{orth}-shift) for the TRRH step is presented which does not control the density change through the overlap of the individual orbitals, but instead controls the amount of new information added to the density subspace. Thus the d^{orth}-shift scheme does not contain any reference to the MO basis and can be used in connection with alternatives to diagonalization. Also, it is found that the d^{orth}-shift scheme leads to a faster convergence since the former level shift scheme is too restrictive, ignoring the well known changes contained in the density subspace.

For TRDSM, an improvement of the energy model is developed, in which a part of the term neglected in the DSM energy model compared to the SCF energy is recovered. However, the effects of the improvement are found rather small compared to the extra complexity added to the algorithm.

An energy minimization algorithm is presented as well, replacing the standard RH-diagonalization in the SCF optimization. The novel idea is to exploit the valuable information saved in the density subspace of the previous densities to construct an improved RH energy model (augmented Roothaan-Hall - ARH) and minimize this model instead of the RH model. This makes the TRDSM step redundant since a density subspace minimization now is included in the minimization of the RH energy model. We expect a faster convergence rate for ARH compared to TRSCF, mainly because the RH and DSM steps are merged to an energy model with correct gradient (not just in the subspace) and an approximate Hessian, which is improved in each iteration using the information from the previous density and Fock matrices. The preliminary results from the ARH energy minimization seems promising, with convergence improvements compared to TRSCF, which already had better or as good convergence rates as DIIS.

The errors introduced in the TRRH and TRDSM energy models compared to the SCF energy are studied. Since the DFT and HF energy expressions differ, the errors in the energy models are potentially different for the two methods. It is found that the DSM energy model has the same error of the order $||\mathbf{D}_\delta||^2$ for both HF and DFT, where $\mathbf{D}_\delta$ is the idempotency correction we impose on the averaged density. For the RH energy model it is found by inspecting test cases that the errors are larger for LDA than for HF, especially when convergence is approached. The error can be divided into two sources, namely the error in the RH Hessian compared to the SCF Hessian, and the size of the third and higher order contributions from the nonlinear terms in the SCF energy, which are not included in the RH energy model. By further tests it seems that the Hessian is better described in LDA than in HF, and since the errors are larger for LDA in particular close to convergence, it seems unlikely that the third and higher order terms are causing the difference. The question why larger errors are seen for LDA than for HF is thus still unanswered and it will be further investigated.

The stability of stationary points is discussed and a method to test and walk away from unstable stationary points is described, and examples are given, where it has been applied. It is acknowledged that such a method is very valuable since otherwise a minimum could not have been found for the examples given.

The scaling of TRSCF is also considered. An alternative to diagonalization has been implemented in our SCF program, where instead of diagonalizing the Fock matrix, the trace purification scheme by Palser and Manolopoulos[19] and later Niklasson[48] is used. The purification scheme in combination with the d^{orth}-shift scheme make the TRRH step near-linearly scaling. The trace purification scheme is linear scaling in an orthogonal basis, but since the optimization scheme is formulated in the non-orthogonal AO basis, the transformation to an orthogonal basis has an N^2 scaling with a small prefactor. Timings for the TRRH step with diagonalizations and with purifications are given, and it

is seen that the trace purification scheme is a major improvement compared to diagonalization when more that a couple of thousand basis functions are needed. The TRDSM step is based on matrix multiplications and additions, so by construction it will be linearly scaling when sparsity in the matrices is exploited.

As illustrated in the examples throughout this part of the thesis and in the applications section, significant improvements to SCF convergence have been obtained. For both the TRSCF and ARH examples presented, the convergence is as good as or better than DIIS, and for problems where DIIS diverges, convergence is obtained with the TRSCF and ARH methods. The globally convergent trust region method by Francisco *et. al.*[26] is found to be better only for the simplest examples whereas for the rest, the TRSCF and ARH methods are found superior. The future success of the TRSCF method depends on a well optimized implementation of the diagonalization alternative combined with the dynamic level shift scheme, and sparsity being exploited in an efficient manner such that it can compete with the linear scaling SCF programs used today. The future success of the ARH method depends on finding efficient ways of solving the nonlinear equations corresponding to the minimization of the energy model. For this purpose different preconditioners will be tested.

To conclude, there are still some adjustments that should be done to improve the algorithms, but the framework is in place. The SCF optimization algorithms presented in this thesis, each make up a black-box optimization scheme for HF and DFT as there is one scheme without any user-adjustment that lead to fast and stable convergence for both simple and problematic systems studied so far. We are thus convinced that TRSCF and ARH are build to handle the optimization problems of the future.

Part 2

Atomic Orbital Based Response Theory

2.1 Introduction

The first part of this thesis was concerned with the optimization of the one electron density matrix for Hartree-Fock (HF) and density-functional theory (DFT). From such an optimized density, information about excited states and how the system reacts to a perturbation (e.g. an external electric field) may be obtained using response theory. Response theory and the derivation of molecular properties will be the subject of this part of the thesis.

Response theory provides a rigorous approach for calculating molecular properties. As for the SCF optimization algorithms, the theory has usually been formulated in the molecular orbital (MO) basis which is inherently delocal, making the implicated matrices non-sparse. A reformulation in the local atomic orbital (AO) basis is thus necessary to obtain linear scaling algorithms and permit calculations of properties for large systems. Such a reformulation, in which an exponential parameterization of the density matrix is employed, is given in a paper by Larsen *et al.*[61].

The AO formulation of the response functions has a number of advantages compared to the MO formulation, besides locality. The response equations and molecular property expressions are simpler in the AO basis as the involved matrices (e.g. the Fock and property matrices) enter the equations in the basis they are evaluated in originally. No transformation between bases is necessary in the AO formulation as it is in the MO formulation. The AO formulation is particular convenient for perturbation dependent basis sets. In the MO formulation a set of perturbation dependent orthonormal molecular orbitals must be introduced. These orbitals have no physical content and thus add artificial complexity to the problem. To exemplify the benefits of the AO formulation, the expression for the excited state geometrical gradient is derived in Section 2.4.

In the conventional MO formulation, number operators are redundant and can be eliminated. However, in the AO basis the number operators are not redundant and must be included. Because of this, the proof of pairing in the solutions of the response equations cannot be directly taken from the MO basis to the AO basis. It is thus necessary to study the impact of the included number operators on the solver for the AO response equations. This has been done in Section 2.2, using the method of second quantization to formulate the AO based response equations. Implementation issues connected to solving the AO response equations are discussed in Section 2.3. In Section 2.5 a couple of simple examples are given, where the AO response solver is used to find ground and excited state properties. In Section 2.6 the results of this part of the thesis are summarized.

2.2 AO Based Response Equations in Second Quantization

In this section the linear response equations are derived for Hartree-Fock theory, but with minor technical changes they apply to DFT as well. The quadratic and higher response equations could equally well be derived in this formulation; however, this is not necessary to arrive at the basic conclusions.

2.2.1 The Parameterization

Consider a set of atomic orbitals (χ_μ) with the real and symmetric metric **S**. The creation and annihilation operators for the atomic orbitals fulfil the anticommutation relation

$$\left[a_\mu^\dagger, a_\nu \right]_+ = S_{\nu\mu} . \tag{2.1}$$

We will consider the following exponential operator

$$\hat{T} = \exp\left(i\hat{\kappa} \right), \tag{2.2}$$

where $\hat{\kappa}$ is a Hermitian one-electron operator

$$\hat{\kappa} = \sum_{\mu\nu} \kappa_{\mu\nu} a_\mu^\dagger a_\nu \tag{2.3}$$

$$\boldsymbol{\kappa}^\dagger = \boldsymbol{\kappa} . \tag{2.4}$$

To examine the action of $\exp(i\hat{\kappa})$, we consider the transformed creation operators

$$\tilde{a}_\mu^\dagger = \exp(i\hat{\kappa}) a_\mu^\dagger \exp(-i\hat{\kappa}) . \tag{2.5}$$

It is seen that the transformed operators satisfy the same anticommutation relations as the untransformed operators

$$\begin{aligned} \left[\tilde{a}_\mu^\dagger, \tilde{a}_\nu \right]_+ &= \left[\exp(i\hat{\kappa}) a_\mu^\dagger \exp(-i\hat{\kappa}), \exp(i\hat{\kappa}) a_\nu \exp(-i\hat{\kappa}) \right]_+ \\ &= \exp(i\hat{\kappa}) \left[a_\mu^\dagger, a_\nu \right]_+ \exp(-i\hat{\kappa}) = S_{\nu\mu} . \end{aligned} \tag{2.6}$$

The exponential operators of Eq. (2.2) are therefore the manifold of operators that conserves the general metric **S**. In the special case where **S** = **1**, the exponential operator reduces to the standard exponential operator occurring in the second quantization formalism of the molecular orbital based method.[46]

Using the Baker-Champbell-Hausdorff expansion[46] and the anticommutation relation of Eq. (2.1), we get

$$\begin{aligned}\tilde{a}_\mu^\dagger &= a_\mu^\dagger + i\left[\hat{\kappa}, a_\mu^\dagger\right] - \tfrac{1}{2}\left[\hat{\kappa},\left[\hat{\kappa}, a_\mu^\dagger\right]\right] + \cdots \\ &= a_\mu^\dagger + i\sum_\nu (\boldsymbol{\kappa}\mathbf{S})_{\nu\mu}\, a_\nu^\dagger - \tfrac{1}{2}\sum_\nu (\boldsymbol{\kappa}\mathbf{S})_{\nu\mu}^2\, a_\nu^\dagger + \cdots . \\ &= \sum_\nu \exp(i\boldsymbol{\kappa}\mathbf{S})_{\nu\mu}\, a_\nu^\dagger .\end{aligned} \tag{2.7}$$

To further investigate the properties of the above exponential transformation, we next consider the transformation of a single determinant state $|0\rangle$ with $\exp(i\hat{\kappa})$

$$\left|\tilde{0}\right\rangle = \exp(i\hat{\kappa})|0\rangle . \tag{2.8}$$

The properties of $\left|\tilde{0}\right\rangle$ may be obtained by comparing the expectation values of transformed creation-annihilation operators

$$\tilde{\Delta}_{\mu\nu} = \left\langle \tilde{0}\left| a_\mu^\dagger a_\nu \right|\tilde{0}\right\rangle = \left\langle 0\left|\exp(-i\hat{\kappa}) a_\mu^\dagger \exp(i\hat{\kappa}) \exp(-i\hat{\kappa}) a_\nu \exp(i\hat{\kappa})\right|0\right\rangle \tag{2.9}$$

with the expectation values of the untransformed operators

$$\Delta_{\mu\nu} = \left\langle 0\left| a_\mu^\dagger a_\nu \right|0\right\rangle . \tag{2.10}$$

To rewrite Eq. (2.9) in terms of Eq. (2.10) we use Eq. (2.7) to write the transformed creation- and annihilation-operators in terms of the untransformed operators

$$\begin{aligned}\exp(-i\hat{\kappa}) a_\mu^\dagger \exp(i\hat{\kappa}) &= \sum_\rho \exp(-i\boldsymbol{\kappa}\mathbf{S})_{\rho\mu}\, a_\rho^\dagger \\ \exp(-i\hat{\kappa}) a_\nu \exp(i\hat{\kappa}) &= \sum_\rho \exp(i\mathbf{S}\boldsymbol{\kappa})_{\nu\rho}\, a_\rho .\end{aligned} \tag{2.11}$$

Substituting these expressions into Eq. (2.9) gives

$$\tilde{\boldsymbol{\Delta}} = \exp\left(-i\mathbf{S}\boldsymbol{\kappa}^{\mathrm{T}}\right)\boldsymbol{\Delta}\exp\left(i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S}\right). \tag{2.12}$$

In Appendix B, it is shown that if $|0\rangle$ is a single determinant wave function, then $\boldsymbol{\Delta}$ fulfils Eqs. (B-7), corresponding to the symmetry, trace, and idempotency condition for the one-electron density. We will now show that if $\boldsymbol{\Delta}$ fulfils these equations then so does $\tilde{\boldsymbol{\Delta}}$. The Hermiticity of $\boldsymbol{\Delta}$ follows from the Hermiticity of **S** and $\boldsymbol{\kappa}$ and will not be shown explicitly here. The trace relation is shown as follows

$$\begin{aligned}\mathrm{Tr}\,\tilde{\mathbf{\Delta}}\mathbf{S}^{-1} &= \mathrm{Tr}\,\mathbf{\Delta}\exp(i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S})\mathbf{S}^{-1}\exp(-i\mathbf{S}\boldsymbol{\kappa}^{\mathrm{T}})\mathbf{S}\mathbf{S}^{-1} \\ &= \mathrm{Tr}\,\mathbf{\Delta}\exp(i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S})\exp(-i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S})\mathbf{S}^{-1} \\ &= \mathrm{Tr}\,\mathbf{\Delta}\mathbf{S}^{-1}\,,\end{aligned} \tag{2.13}$$

where we have used the relation

$$\mathbf{B}^{-1}\exp(\mathbf{A})\mathbf{B} = \exp(\mathbf{B}^{-1}\mathbf{A}\mathbf{B})\,. \tag{2.14}$$

The same relation may be used to show the idempotency relation

$$\begin{aligned}\tilde{\mathbf{\Delta}}\mathbf{S}^{-1}\tilde{\mathbf{\Delta}} &= \exp\left(-i\mathbf{S}\boldsymbol{\kappa}^{\mathrm{T}}\right)\mathbf{\Delta}\exp\left(i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S}\right)\mathbf{S}^{-1}\exp\left(-i\mathbf{S}\boldsymbol{\kappa}^{\mathrm{T}}\right)\mathbf{\Delta}\exp\left(i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S}\right) \\ &= \exp\left(-i\mathbf{S}\boldsymbol{\kappa}^{\mathrm{T}}\right)\mathbf{\Delta}\exp\left(i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S}\right)\exp\left(-i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S}\right)\mathbf{S}^{-1}\mathbf{\Delta}\exp\left(i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S}\right) \\ &= \exp\left(-i\mathbf{S}\boldsymbol{\kappa}^{\mathrm{T}}\right)\mathbf{\Delta}\mathbf{S}^{-1}\mathbf{\Delta}\exp\left(i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S}\right) \\ &= \exp\left(-i\mathbf{S}\boldsymbol{\kappa}^{\mathrm{T}}\right)\mathbf{\Delta}\exp\left(i\boldsymbol{\kappa}^{\mathrm{T}}\mathbf{S}\right) = \tilde{\mathbf{\Delta}}\,.\end{aligned} \tag{2.15}$$

We can therefore conclude that $\tilde{\mathbf{\Delta}}$ fulfils Eqs. (B-7) and $\exp(i\hat{\kappa})|0\rangle$ is therefore a legitimate normalized single-determinant wave function. It can be shown that all matrices fulfilling Eqs. (B-7) can be obtained from an appropriate choice of $\boldsymbol{\kappa}$, so the transformation of Eq. (2.8) is a complete parameterization.

2.2.2 The Linear Response Function

We will now use the parameterization of Eq. (2.8) for an arbitrary single-determinant wave function to describe a Hartree-Fock wave function in an external, time-dependent field. The parameters in $\boldsymbol{\kappa}$ will become time-dependent and we will in the following develop equations for obtaining these parameters. The time-dependent Hamiltonian can be written as

$$H = H_0 + V_t\,, \tag{2.16}$$

where H_0 is the Hamiltonian for the unperturbed system, and V_t is a first-order perturbation. The perturbation will be turned on adiabatically, and V_t can be expressed as

$$V_t = \int_{-\infty}^{\infty} \mathrm{d}\omega V_\omega \exp((-i\omega + \varepsilon)t)\,, \tag{2.17}$$

where ε is a positive infinitesimal that ensures $V_t \to 0$ as $t \to -\infty$. The perturbation is required to be Hermitian, so we have the relation

$$V_\omega^\dagger = V_{-\omega}\,. \tag{2.18}$$

To determine the linear response function, we begin by considering the time dependence of the expectation value $\langle\tilde{0}|A|\tilde{0}\rangle$ of a one-electron operator A. We need only expand the wave function $|\tilde{0}\rangle$ of Eq. (2.8) to first order in the external perturbation to obtain the linear response:

$$\hat{\kappa} = \hat{\kappa}_t^{(1)} + \hat{\kappa}_t^{(2)} + \cdots\,. \tag{2.19}$$

The zero-order contribution, $\hat{\kappa}_t^{(0)}$, vanishes as the unperturbed wave function $|0\rangle$ is assumed to be optimized for the zero-order Hamiltonian, so the Brillouin-conditions in the AO basis hold

$$\frac{\partial}{\partial \kappa_{\mu\nu}} \langle 0|H_0|0\rangle = i\left\langle 0\middle|\left[H_0, a_\mu^\dagger a_\nu \right]\middle|0\right\rangle = 0 . \tag{2.20}$$

Substitution of the expansion of $\hat{\kappa}$ into Eq. (2.8) gives to first order:

$$\left\langle \tilde{0}|A|\tilde{0}\right\rangle = \langle 0|A|0\rangle - i\left\langle 0\middle|\left[\hat{\kappa}_t^{(1)}, A \right]\middle|0\right\rangle . \tag{2.21}$$

Since the response functions are defined in the frequency rather than the time domain, we formulate the wave function corrections in the frequency space. By analogy with Eq. (2.17), we write

$$\kappa_t^{(1)} = \int_{-\infty}^{\infty} \mathrm{d}\omega\, \kappa_\omega^{(1)} \exp\left((-i\omega + \varepsilon)\, t \right) . \tag{2.22}$$

Inserting Eq. (2.22) into Eq. (2.21) we obtain

$$\left\langle \tilde{0}|A|\tilde{0}\right\rangle = \langle 0|A|0\rangle - i\int_{-\infty}^{\infty} \mathrm{d}\omega \left\langle 0\middle|\left[\hat{\kappa}_\omega^{(1)}, A \right]\middle|0\right\rangle \exp\left((-i\omega + \varepsilon)t \right) . \tag{2.23}$$

Comparing Eq. (2.23) with the formal expansion of an expectation value in terms of a response function

$$\left\langle \tilde{0}|A|\tilde{0}\right\rangle = \langle 0|A|0\rangle + \int_{-\infty}^{\infty} \mathrm{d}\omega \langle\langle A; V_\omega \rangle\rangle_\omega \exp\left((-i\omega + \varepsilon)t \right) , \tag{2.24}$$

we may identify the linear response function as

$$\langle\langle A; V_\omega \rangle\rangle_\omega = -i\left\langle 0\middle|\left[\hat{\kappa}_\omega^{(1)}, A \right]\middle|0\right\rangle . \tag{2.25}$$

2.2.3 The Time Development of the Reference State

Before the explicit time-dependent equations are set up for determining the time-dependent parameters of $\boldsymbol{\kappa}$, it is convenient to rewrite $\hat{\kappa}$, Eq. (2.3), as

$$\hat{\kappa} = \sum_{\mu>\nu} \left(\kappa_{\mu\nu} a_\mu^\dagger a_\nu + \kappa_{\mu\nu}^* a_\nu^\dagger a_\mu \right) + \sum_\mu \kappa_{\mu\mu} a_\mu^\dagger a_\mu , \tag{2.26}$$

which follows from the Hermiticity of $\hat{\kappa}$. The operators of $\hat{\kappa}$ may be collected in a vector (here in row form):

$$\boldsymbol{\Lambda} = \left(\mathbf{Q}^\dagger \quad \mathbf{D}^\dagger \quad \mathbf{Q} \right), \tag{2.27}$$

where the three classes of operators are defined as

$$\begin{aligned} Q_m^\dagger &= a_\mu^\dagger a_\nu \ , \mu > \nu \\ D_m^\dagger &= a_\mu^\dagger a_\mu \\ Q_m &= a_\nu^\dagger a_\mu \ , \mu > \nu . \end{aligned} \tag{2.28}$$

The parameters of $\boldsymbol{\kappa}$ may similarly be arranged in a vector

$$\boldsymbol{\alpha}^{(i)} = \begin{pmatrix} \kappa_{\mu\nu}^{(i)} & \mu > \nu \\ \kappa_{\mu\mu}^{(i)} & \\ \kappa_{\mu\nu}^{(i)*} & \mu > \nu \end{pmatrix}, \tag{2.29}$$

such that

$$\hat{\kappa}^{(i)} = \sum_m \alpha_m^{(i)} \Lambda_m \,. \tag{2.30}$$

Here the index m on Λ runs over all three classes of operators listed in Eq. (2.28).

The single excitation operators $a_\mu^\dagger a_\nu$ have by Eq. (2.27)-(2.28) been divided into a set of atomic orbital excitations, corresponding to $\mu > \nu$ and a set of atomic orbital deexcitations, corresponding to $\mu < \nu$. As the atomic orbital excitations and deexcitation have the same formal properties, this division does not have any physical content. However, the division will prove important when the paired structure of the response equations is investigated in Section 2.2.5. Note that it is not possible to exclude the number operators $a_\mu^\dagger a_\mu$ in the atomic orbital representation, whereas they are redundant in the standard molecular orbital formulation.

In the presence of the time-dependent perturbation, we introduce the time transformed operator basis

$$\tilde{\boldsymbol{\Lambda}}^\dagger = \begin{pmatrix} \tilde{\mathbf{Q}} \\ \tilde{\mathbf{D}} \\ \tilde{\mathbf{Q}}^\dagger \end{pmatrix}, \tag{2.31}$$

where

$$\tilde{Q}_m = \exp(i\hat{\kappa}) Q_m \exp(-i\hat{\kappa}) \tag{2.32}$$

and similarly for $\tilde{Q}_m^\dagger$ and $\tilde{D}_m$.

The time evolution of $\left|\tilde{0}\right\rangle$ may now be determined using Ehrenfest's theorem for the transformed operators of $\tilde{\boldsymbol{\Lambda}}^\dagger$ in Eq. (2.31):

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle \tilde{0}\middle|\tilde{\boldsymbol{\Lambda}}^\dagger\middle|\tilde{0}\right\rangle - \left\langle \tilde{0}\middle|\left(\frac{\partial}{\partial t}\tilde{\boldsymbol{\Lambda}}^\dagger\right)\middle|\tilde{0}\right\rangle = -i\left\langle \tilde{0}\middle|\left[\tilde{\boldsymbol{\Lambda}}^\dagger, H_0 + V_t\right]\middle|\tilde{0}\right\rangle. \tag{2.33}$$

2.2.4 The First-order Equation

We now expand Eq. (2.33) in orders of the external perturbation, restricting ourselves to terms that are linear in the amplitudes. Inserting Eq. (2.19) into Eq. (2.33) and collecting the terms linear in the perturbation, we obtain the first-order time-dependent equation

$$i\left\langle 0\left|\left[\boldsymbol{\Lambda}^{\dagger},\dot{\hat{\kappa}}_t^{(1)}\right]\right|0\right\rangle = -i\left\langle 0\left|\left[\boldsymbol{\Lambda}^{\dagger},V_t\right]\right|0\right\rangle + \left\langle 0\left|\left[\boldsymbol{\Lambda}^{\dagger},\left[H_0,\hat{\kappa}_t^{(1)}\right]\right]\right|0\right\rangle. \tag{2.34}$$

To solve the time-dependent equation Eq. (2.34), we insert the frequency expansion of the wave function correction of Eq. (2.22) and of the external perturbation Eq. (2.17)

$$\begin{aligned}&\int_{-\infty}^{\infty} \mathrm{d}\omega \exp\left((-i\omega+\varepsilon)t\right)\left(\omega\left\langle 0\left|\left[\boldsymbol{\Lambda}^{\dagger},\hat{\kappa}_{\omega}^{(1)}\right]\right|0\right\rangle - \left\langle 0\left|\left[\boldsymbol{\Lambda}^{\dagger},\left[H_0,\hat{\kappa}_{\omega}^{(1)}\right]\right]\right|0\right\rangle\right)\\ &= \int_{-\infty}^{\infty} \mathrm{d}\omega \exp\left((-i\omega+\varepsilon)t\right)\left(-i\left\langle 0\left|\left[\boldsymbol{\Lambda}^{\dagger},V_{\omega}\right]\right|0\right\rangle\right) \quad .\end{aligned} \tag{2.35}$$

The first-order response equation is then found as

$$\omega\left\langle 0\left|\left[\boldsymbol{\Lambda}^{\dagger},\hat{\kappa}_{\omega}^{(1)}\right]\right|0\right\rangle - \left\langle 0\left|\left[\boldsymbol{\Lambda}^{\dagger},\left[H_0,\hat{\kappa}_{\omega}^{(1)}\right]\right]\right|0\right\rangle = -i\left\langle 0\left|\left[\boldsymbol{\Lambda}^{\dagger},V_{\omega}\right]\right|0\right\rangle. \tag{2.36}$$

The equation may be written in terms of the matrices

$$E_{mn}^{[2]} = \left\langle 0\left|\left[\Lambda_m^{\dagger},[H_0,\Lambda_n]\right]\right|0\right\rangle, \tag{2.37}$$

$$S_{mn}^{[2]} = \left\langle 0\left|\left[\Lambda_m^{\dagger},\Lambda_n\right]\right|0\right\rangle, \tag{2.38}$$

and the vector

$$\left[V_{\omega}^{[1]}\right]_m = \left\langle 0\left|\left[\Lambda_m^{\dagger},V_{\omega}\right]\right|0\right\rangle. \tag{2.39}$$

Using Eqs. (2.37)-(2.39) and (2.29)-(2.30), we now write the first-order response equations, Eq. (2.36), in the form

$$\left(\mathbf{E}^{[2]} - \omega\mathbf{S}^{[2]}\right)\boldsymbol{\alpha}^{(1)} = i\mathbf{V}_{\omega}^{[1]}, \tag{2.40}$$

where $\mathbf{E}^{[2]}$ and $\mathbf{S}^{[2]}$ may be viewed as generalized electronic Hessian and overlap matrices[61,62]. The matrix elements $E_{mn}^{[2]}$ and $S_{mn}^{[2]}$ (Eq. (2.37) and (2.38)) can be expressed as matrix multiplications and additions of the density, Fock and overlap matrices.[61]

The linear response function is obtained by inserting the first-order correction as obtained in Eq. (2.40) in the expression for the linear response function Eq. (2.25). Renaming the perturbation operator V_{ω} to B and introducing

$$\begin{aligned}A_m^{[1]} &= -\left\langle 0\left|\left[\Lambda_m, A\right]\right|0\right\rangle\\ B_m^{[1]} &= \left\langle 0\left|\left[\Lambda_m^{\dagger}, B\right]\right|0\right\rangle\end{aligned} \tag{2.41}$$

we obtain

$$\left\langle\left\langle A;B\right\rangle\right\rangle_{\omega} = -\mathbf{A}^{[1]}\left(\mathbf{E}^{[2]} - \omega\mathbf{S}^{[2]}\right)^{-1}\mathbf{B}^{[1]}. \tag{2.42}$$

The linear response function may thus be calculated by solving one set of linear equations at each frequency. To be more explicit, denoting the solution vector to the linear response equation

$$\mathbf{N}^{B}(\omega) = \left(\mathbf{E}^{[2]} - \omega\mathbf{S}^{[2]}\right)^{-1}\mathbf{B}^{[1]}, \tag{2.43}$$

the linear response function in Eq. (2.42) can be obtained as

$$\langle\langle A;B\rangle\rangle_{\omega} = -\mathbf{A}^{[1]}\mathbf{N}^{B}(\omega)\,. \tag{2.44}$$

2.2.5 Pairing

The excitation energies are identified as the poles of the linear response function of Eq. (2.42) and are therefore solutions to the generalized eigenvalue problem

$$\mathbf{E}^{[2]}\mathbf{X} = \omega\mathbf{S}^{[2]}\mathbf{X}\,. \tag{2.45}$$

In the MO formulation of response theory, it has been shown that the excitation energies are paired[63], so that if ω_i is an eigenvalue for Eq. (2.45) then so is $-\omega_i$. It is important to understand how pairing appears in the AO basis, in particular since this structural feature is exploited when the equations are solved iteratively as is necessary for large problems. This is further discussed in Section 2.3. Since the proof of the pairing given in the MO formulation cannot be directly transferred to the AO formulation due to the presence of the diagonal operators D_m, this section gives the proof in the AO formulation.

The structure of $\mathbf{E}^{[2]}$ and $\mathbf{S}^{[2]}$ in the AO formulation is analyzed for the purpose of examining the pairing structure. Dividing $\mathbf{\Lambda}$ into the tree classes of Eq. (2.28), the matrix $\mathbf{E}^{[2]}$ may be written as

$$\mathbf{E}^{[2]} = \begin{pmatrix} \langle 0|\big[\mathbf{Q},\big[H_0,\mathbf{Q}^{\dagger}\big]\big]|0\rangle & \langle 0|[\mathbf{Q},[H_0,\mathbf{D}]]|0\rangle & \langle 0|[\mathbf{Q},[H_0,\mathbf{Q}]]|0\rangle \\ \langle 0|\big[\mathbf{D},\big[H_0,\mathbf{Q}^{\dagger}\big]\big]|0\rangle & \langle 0|[\mathbf{D},[H_0,\mathbf{D}]]|0\rangle & \langle 0|[\mathbf{D},[H_0,\mathbf{Q}]]|0\rangle \\ \langle 0|\big[\mathbf{Q}^{\dagger},\big[H_0,\mathbf{Q}^{\dagger}\big]\big]|0\rangle & \langle 0|\big[\mathbf{Q}^{\dagger},[H_0,\mathbf{D}]\big]|0\rangle & \langle 0|\big[\mathbf{Q}^{\dagger},[H_0,\mathbf{Q}]\big]|0\rangle \end{pmatrix}. \tag{2.46}$$

If we assume for simplicity that all orbitals and integrals for the unperturbed system are real, the elements of for example the block $\langle 0|\big[\mathbf{Q}^{\dagger},[H_0,\mathbf{Q}]\big]|0\rangle$ are trivially rewritten as

$$\begin{aligned} \langle 0|\big[Q_m^{\dagger},[H_0,Q_n]\big]|0\rangle &= \langle 0|\big[Q_m^{\dagger},[H_0,Q_n]\big]|0\rangle^{*} \\ &= \langle 0|\big[Q_m,\big[H_0,Q_n^{\dagger}\big]\big]|0\rangle\,. \end{aligned} \tag{2.47}$$

The nine blocks in Eq. (2.46) can then all be written in terms of the following four matrices

$$\begin{aligned} A_{mn} &= \langle 0|\big[Q_m,\big[H_0,Q_n^{\dagger}\big]\big]|0\rangle\,, \\ B_{mn} &= \langle 0|\big[Q_m,\big[H_0,Q_n\big]\big]|0\rangle\,, \\ F_{mn} &= \langle 0|\big[Q_m,\big[H_0,D_n\big]\big]|0\rangle\,, \\ G_{mn} &= \langle 0|\big[D_m,\big[H_0,D_n\big]\big]|0\rangle\,, \end{aligned} \tag{2.48}$$

and we obtain

$$\mathbf{E}^{[2]} = \begin{pmatrix} \mathbf{A} & \mathbf{F} & \mathbf{B} \\ \mathbf{F}^{\mathrm{T}} & \mathbf{G} & \mathbf{F}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{F} & \mathbf{A} \end{pmatrix}. \tag{2.49}$$

The matrix $\mathbf{S}^{[2]}$ may in a similar way be written as

$$\mathbf{S}^{[2]} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{\Omega} & \mathbf{\Delta} \\ \mathbf{\Omega}^{\mathrm{T}} & \mathbf{0} & -\mathbf{\Omega}^{\mathrm{T}} \\ -\mathbf{\Delta} & -\mathbf{\Omega} & -\mathbf{\Sigma} \end{pmatrix}, \tag{2.50}$$

where

$$\begin{aligned} \Sigma_{mn} &= \langle 0 | [\, Q_m, Q_n^{\dagger} \,] | 0 \rangle, \\ \Delta_{mn} &= \langle 0 | [\, Q_m, Q_n \,] | 0 \rangle, \\ \Omega_{mn} &= \langle 0 | [\, Q_m, D_n \,] | 0 \rangle. \end{aligned} \tag{2.51}$$

Note that the block containing two diagonal operators vanishes as

$$\langle 0 | [\, D_m, D_n \,] | 0 \rangle = \langle 0 | [\, a_\mu^{\dagger} a_\mu, a_\nu^{\dagger} a_\nu \,] | 0 \rangle = S_{\mu\nu} \langle 0 | a_\mu^{\dagger} a_\nu | 0 \rangle - S_{\nu\mu} \langle 0 | a_\nu^{\dagger} a_\mu | 0 \rangle = 0 . \tag{2.52}$$

To illustrate how the pairing is obtained in the AO formulation, we assume that the vector

$$\mathbf{X} = \begin{pmatrix} \mathbf{Z} \\ \mathbf{U} \\ \mathbf{Y} \end{pmatrix} \tag{2.53}$$

is an eigenvector for Eq. (2.45) with eigenvalue ω

$$\begin{pmatrix} \mathbf{A} & \mathbf{F} & \mathbf{B} \\ \mathbf{F}^{\mathrm{T}} & \mathbf{G} & \mathbf{F}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{F} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{Z} \\ \mathbf{U} \\ \mathbf{Y} \end{pmatrix} = \omega \begin{pmatrix} \mathbf{\Sigma} & \mathbf{\Omega} & \mathbf{\Delta} \\ \mathbf{\Omega}^{\mathrm{T}} & \mathbf{0} & -\mathbf{\Omega}^{\mathrm{T}} \\ -\mathbf{\Delta} & -\mathbf{\Omega} & -\mathbf{\Sigma} \end{pmatrix} \begin{pmatrix} \mathbf{Z} \\ \mathbf{U} \\ \mathbf{Y} \end{pmatrix}. \tag{2.54}$$

Multiplying the blocks of Eq. (2.54) gives three sets of equations

$$\begin{aligned} \mathbf{AZ} + \mathbf{FU} + \mathbf{BY} &= \omega (\mathbf{\Sigma Z} + \mathbf{\Omega U} + \mathbf{\Delta Y}) \\ \mathbf{F}^{\mathrm{T}}\mathbf{Z} + \mathbf{GU} + \mathbf{F}^{\mathrm{T}}\mathbf{Y} &= \omega \left(\mathbf{\Omega}^{\mathrm{T}}\mathbf{Z} - \mathbf{\Omega}^{\mathrm{T}}\mathbf{Y} \right) \\ \mathbf{BZ} + \mathbf{FU} + \mathbf{AY} &= \omega (-\mathbf{\Delta Z} - \mathbf{\Omega U} - \mathbf{\Sigma Y}) . \end{aligned} \tag{2.55}$$

We will now prove that the paired vector

$$\mathbf{X}^{P} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{U} \\ \mathbf{Z} \end{pmatrix} \tag{2.56}$$

is an eigenvector for Eq. (2.45) with eigenvalue $-\omega$

$$\begin{pmatrix} \mathbf{A} & \mathbf{F} & \mathbf{B} \\ \mathbf{F}^{\mathrm{T}} & \mathbf{G} & \mathbf{F}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{F} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{Y} \\ \mathbf{U} \\ \mathbf{Z} \end{pmatrix} = -\omega \begin{pmatrix} \mathbf{\Sigma} & \mathbf{\Omega} & \mathbf{\Delta} \\ \mathbf{\Omega}^{\mathrm{T}} & \mathbf{0} & -\mathbf{\Omega}^{\mathrm{T}} \\ -\mathbf{\Delta} & -\mathbf{\Omega} & -\mathbf{\Sigma} \end{pmatrix} \begin{pmatrix} \mathbf{Y} \\ \mathbf{U} \\ \mathbf{Z} \end{pmatrix}. \tag{2.57}$$

Multiplying the blocks of Eq. (2.57) leads to the three sets of equations

$$\begin{aligned}
\mathbf{AY} + \mathbf{FU} + \mathbf{BZ} &= -\omega(\mathbf{\Sigma Y} + \mathbf{\Omega U} + \mathbf{\Delta Z}) \\
\mathbf{F}^{\mathrm{T}}\mathbf{Y} + \mathbf{GU} + \mathbf{F}^{\mathrm{T}}\mathbf{Z} &= -\omega\left(\mathbf{\Omega}^{\mathrm{T}}\mathbf{Y} - \mathbf{\Omega}^{\mathrm{T}}\mathbf{Z}\right) \\
\mathbf{BY} + \mathbf{FU} + \mathbf{AZ} &= -\omega(-\mathbf{\Delta Y} - \mathbf{\Omega U} - \mathbf{\Sigma Z}),
\end{aligned} \tag{2.58}$$

which are identical to Eqs. (2.55). It is thus concluded that if $\mathbf{X}$ is an eigenvector of Eq. (2.45) with eigenvalue ω, then $\mathbf{X}^P$ is also an eigenvector with eigenvalue $-\omega$.

2.3 Solving the Response Equations

For large systems, the response equations

$$\left(\mathbf{E}^{[2]} - \omega\mathbf{S}^{[2]}\right)\mathbf{N}^B(\omega) = \mathbf{B}^{[1]} \tag{2.59}$$

are best solved using iterative algorithms. These algorithms rely on the ability to set up linear transformations. Expressions for $\mathbf{E}^{[2]}\mathbf{b}$ and $\mathbf{S}^{[2]}\mathbf{b}$, where $\mathbf{b}$ is a trial vector, have previously been derived.[61]

$$\boldsymbol{\sigma} = \mathbf{E}^{[2]}\mathbf{b} \tag{2.60}$$

$$\boldsymbol{\rho} = \mathbf{S}^{[2]}\mathbf{b}\,. \tag{2.61}$$

In each iteration, the response equations are set up and solved in a reduced space. For a reduced space consisting of k trial vectors, the equations can be written as

$$\left(\mathbf{E}^{[2]}_{\mathrm{RED}} - \omega\mathbf{S}^{[2]}_{\mathrm{RED}}\right)\mathbf{X}^{\mathrm{RED}} = \mathbf{B}^{[1]}_{\mathrm{RED}}\,, \tag{2.62}$$

where the reduced matrices are found as

$$\begin{aligned}
\left[\,\mathbf{E}^{[2]}_{\mathrm{RED}}\,\right]_{ij} &= \mathbf{b}_i^{\mathrm{T}}\mathbf{E}^{[2]}\mathbf{b}_j = \mathbf{b}_i^{\mathrm{T}}\boldsymbol{\sigma}_j \\
\left[\,\mathbf{S}^{[2]}_{\mathrm{RED}}\,\right]_{ij} &= \mathbf{b}_i^{\mathrm{T}}\mathbf{S}^{[2]}\mathbf{b}_j = \mathbf{b}_i^{\mathrm{T}}\boldsymbol{\rho}_j \\
\left[\,\mathbf{B}^{[1]}_{\mathrm{RED}}\,\right]_{i} &= \mathbf{b}_i^{\mathrm{T}}\mathbf{B}^{[1]}\,.
\end{aligned} \tag{2.63}$$

Normally when this type of iterative procedure is used, the reduced space is extended with one new trial vector in each iteration. However, due to the pairing described in the previous section, the linear transformations of $\mathbf{E}^{[2]}$ and $\mathbf{S}^{[2]}$ on a trial vector, here exemplified by $\mathbf{E}^{[2]}\mathbf{b}$,

$$\mathbf{E}^{[2]}\mathbf{b} = \begin{pmatrix} \mathbf{A} & \mathbf{F} & \mathbf{B} \\ \mathbf{F}^{\mathrm{T}} & \mathbf{G} & \mathbf{F}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{F} & \mathbf{A} \end{pmatrix}\begin{pmatrix} \mathbf{Z} \\ \mathbf{U} \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} \mathbf{AZ} + \mathbf{FU} + \mathbf{BY} \\ \mathbf{F}^{\mathrm{T}}\mathbf{Z} + \mathbf{GU} + \mathbf{F}^{\mathrm{T}}\mathbf{Y} \\ \mathbf{BZ} + \mathbf{FU} + \mathbf{AY} \end{pmatrix} = \boldsymbol{\sigma}\,, \tag{2.64}$$

may be obtained directly for the paired trial vector as well

$$\mathbf{E}^{[2]}\mathbf{b}^P = \begin{pmatrix} \mathbf{A} & \mathbf{F} & \mathbf{B} \\ \mathbf{F}^{\mathrm{T}} & \mathbf{G} & \mathbf{F}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{F} & \mathbf{A} \end{pmatrix}\begin{pmatrix} \mathbf{Y} \\ \mathbf{U} \\ \mathbf{Z} \end{pmatrix} = \begin{pmatrix} \mathbf{AY} + \mathbf{FU} + \mathbf{BZ} \\ \mathbf{F}^{\mathrm{T}}\mathbf{Y} + \mathbf{GU} + \mathbf{F}^{\mathrm{T}}\mathbf{Z} \\ \mathbf{BY} + \mathbf{FU} + \mathbf{AZ} \end{pmatrix} = \boldsymbol{\sigma}^P\,. \tag{2.65}$$

The reduced space is therefore extended with both vectors without additional cost. Furthermore, when a trial vector and its paired counterpart are simultaneously added to the reduced space, the paired structure of the response equations is preserved. With this structure preserved, the eigenvalues in the reduced space will also be real and paired, and the lowest eigenvalue will monotonically decrease towards the converged value as the reduced space is increased.[64]

The solution vector in the reduced space $\mathbf{X}^{\mathrm{RED}}$, can be expanded in the basis of trial vectors to express the solution vector in the full space

$$\tilde{\mathbf{N}}^{B} = \sum_{i=1}^{k} \left(X_i^{\mathrm{RED}} \mathbf{b}_i \right). \tag{2.66}$$

The residual can then be found as

$$\begin{aligned} \mathbf{R}_k &= \left(\mathbf{E}^{[2]} - \omega \mathbf{S}^{[2]} \right) \tilde{\mathbf{N}}^{B} - \mathbf{B}^{[1]} \\ &= \sum_{i=1}^{k} X_i^{\mathrm{RED}} \left(\boldsymbol{\sigma}_i - \omega \boldsymbol{\rho}_i \right) - \mathbf{B}^{[1]} . \end{aligned} \tag{2.67}$$

If the norm of the residual is smaller than some specified tolerance, the iterative procedure is ended and the converged solution vector has been found

$$\mathbf{N}^{B}(\omega) = \tilde{\mathbf{N}}^{B} . \tag{2.68}$$

If the residual is too large, a new trial vector may be generated from the residual, preferably with a preconditioner $\mathbf{A}$ to speed up the convergence

$$\mathbf{b}_{k+1} = \mathbf{A}^{-1} \mathbf{R}_k . \tag{2.69}$$

The reduced space is then extended with $\mathbf{b}_{k+1}$ and $\mathbf{b}_{k+2} = \mathbf{b}_{k+1}^{P}$ and Eq. (2.62) is set up and solved again, establishing the iterative procedure.

2.3.1 Preconditioning

As mentioned above, the residual found in each iteration should be preconditioned to obtain an effective solver. As a consequence of the strict AO formulation, the electronic Hessian has no diagonal dominance as was the case in the MO basis. This makes preconditioning a challenge. So far, this problem has not been solved in our SCF response solver. Instead, a transformation is made to the MO basis, where the preconditioning is carried out in the usual way using the orbital eigenvalue differences,

$$\left[\mathbf{b}_{k+1}^{\mathrm{MO}} \right]_{ai} = \left[\mathbf{C}^{\mathrm{T}} \mathbf{R}_k \mathbf{C} \right]_{ai} \Big/ \left(\varepsilon_a - \varepsilon_i \right), \tag{2.70}$$

where **C** is the MO expansion coefficients and ε the orbital energies of the reference state. The index *a* refers to virtual orbitals and *i* refers to occupied orbitals. The resulting vector is then back transformed to the AO basis

$$\mathbf{b}_{k+1} = \mathbf{C}\mathbf{b}_{k+1}^{\mathrm{MO}}\mathbf{C}^{\mathrm{T}} . \tag{2.71}$$

An AO alternative to this preconditioner should of course be found, since the reference to the MO basis in this preconditioner introduces dense matrix intermediates. Moreover, at least one diagonalization should be carried out at the end of the optimization of the reference state to obtain the information on the MOs.

2.3.2 Projections

In the MO basis, the orbital rotations within the occupied and virtual spaces are redundant. The response equations in the MO formulation are thus simply set up in the non-redundant occupied-virtual space to avoid linear dependencies. In the AO basis no such separation exists and the equations are set up in the full space. To avoid redundancies in the AO formulation, projections onto the non-redundant space should be made. In the exponential parameterization of the density matrix used in our AO formulation of the response functions, the projector[23]

$$\begin{aligned} \mathcal{P} &= \mathbf{P}\otimes\mathbf{Q} + \mathbf{Q}\otimes\mathbf{P} \\ (\mathcal{P}\mathbf{X})_{\mu\nu} &= \sum_{\rho\sigma}\mathcal{P}_{\mu\nu,\rho\sigma}X_{\rho\sigma} = \left(\mathbf{P}\mathbf{X}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}\mathbf{X}\mathbf{P}^{\mathrm{T}}\right)_{\mu\nu} , \end{aligned} \tag{2.72}$$

where

$$\begin{aligned} \mathbf{P} &= \mathbf{D}\mathbf{S} \\ \mathbf{Q} &= \mathbf{1} - \mathbf{D}\mathbf{S} , \end{aligned} \tag{2.73}$$

projects onto the non-redundant parameter space. It can be shown that all new trial vectors **b** and linear transformations **σ** and **ρ** should be projected onto the non-redundant space in the following manner

$$\begin{aligned} \tilde{\mathbf{b}}_{k+1} &= \mathcal{P}\mathbf{b}_{k+1} , \\ \tilde{\boldsymbol{\sigma}}_{k+1} &= \mathcal{P}^{\mathrm{T}}\boldsymbol{\sigma}_{k+1} , \\ \tilde{\boldsymbol{\rho}}_{k+1} &= \mathcal{P}^{\mathrm{T}}\boldsymbol{\rho}_{k+1} . \end{aligned} \tag{2.74}$$

When solving the response equations as described in the beginning of this section, the vectors projected as in Eq. (2.74) are used.

2.4 The Excited State Gradient

In this section the expression for the geometrical gradient of the singlet excited state is derived, to illustrate how expressions for properties can straightforwardly be derived in the AO response framework.

As for the derivations in Section 2.2 we assume that the wave function of the ground state is optimized at the point of the potential surface, $\mathbf{x}_0$, where the excited state gradient is evaluated. The variational condition is thus fulfilled at that point

$$\mathbf{FDS} - \mathbf{SDF} = 0\,, \tag{2.75}$$

and the ground-state energy at $\mathbf{x}_0$ is further obtained as

$$E^0 = 2\,\mathrm{Tr}\,\mathbf{hD} + \mathrm{Tr}\,\mathbf{DG}(\mathbf{D}) + h_{\mathrm{nuc}}\,, \tag{2.76}$$

where **h** is the one-electron Hamiltonian matrix in the AO basis, h_{nuc} is the nuclear-nuclear repulsion, **G** holds the two-electron AO integrals and the Fock matrix **F** is given by **h** + **G**(**D**).

As mentioned previously, the excitation energy corresponding to the excitation from the ground state $|0\rangle$ to the excited state $|f\rangle$ can be found from the poles of the linear response function for the optimized ground state,[62] i.e. as the eigenvalue of the linear response generalized eigenvalue equation as Eq. (2.45)

$$\left(\mathbf{E}^{[2]} - \omega_f \mathbf{S}^{[2]}\right)\mathbf{b}^f = 0\,, \tag{2.77}$$

where ω_f is the electronic excitation energy

$$\omega_f = E^f - E^0 \tag{2.78}$$

and $\mathbf{b}^f$ is the normalized eigenvector.[61,62]

The excitation energy can then be obtained from Eq. (2.77) as

$$\omega_f = \mathbf{b}^{f\dagger}\mathbf{E}^{[2]}\mathbf{b}^f\,, \tag{2.79}$$

assuming that the eigenvectors $\mathbf{b}^f$ satisfy the normalization condition

$$\mathbf{b}^{f\dagger}\mathbf{S}^{[2]}\mathbf{b}^f = \mathbf{1}\,. \tag{2.80}$$

Since we are interested in the molecular gradient for the excited state, $|f\rangle$, the energy of the excited state should be defined at arbitrary points on the potential surface.

2.4.1 Construction of the Lagrangian

The analytic expression for the excited state gradient is found using the Lagrangian technique[65]. We construct the Lagrangian for the excited state energy $E^f = E^0 + \omega_f$, using a matrix-vector notation,

$$L^f = E^0 + \mathbf{b}^{f\dagger}\mathbf{E}^{[2]}\mathbf{b}^f - \bar{\omega}\left(\mathbf{b}^{f\dagger}\mathbf{S}^{[2]}\mathbf{b}^f - 1\right) - \bar{\mathbf{X}}^{\dagger}(\mathbf{FDS} - \mathbf{SDF})\,. \tag{2.81}$$

The variational condition on the ground state, Eq. (2.75), and the orthonormality constraint condition on the eigenvectors, Eq. (2.80), are included, and they are multiplied by the Lagrange multipliers $\bar{\omega}$ and $\bar{\mathbf{X}}$, respectively.

We then require the Lagrangian to be variational in all parameters

$$\frac{\partial L^f}{\partial \bar{\mathbf{X}}} = \mathbf{SDF} - \mathbf{FDS} = 0 \tag{2.82}$$

$$\frac{\partial L^f}{\partial \bar{\omega}} = \mathbf{b}^{f\dagger}\mathbf{S}^{[2]}\mathbf{b}^f - 1 = 0 \tag{2.83}$$

$$\frac{\partial L^f}{\partial \mathbf{b}^{f\dagger}} = \mathbf{E}^{[2]}\mathbf{b}^f - \bar{\omega}\mathbf{S}^{[2]}\mathbf{b}^f = 0 \tag{2.84}$$

$$\frac{\partial L^f}{\partial \mathbf{b}^f} = \mathbf{b}^{f\dagger}\mathbf{E}^{[2]} - \bar{\omega}\mathbf{b}^{f\dagger}\mathbf{S}^{[2]} = 0 \tag{2.85}$$

$$\frac{\partial L^f}{\partial X_m} = \frac{\partial E^0}{\partial X_m} + \frac{\partial \mathbf{b}^{f\dagger}\mathbf{E}^{[2]}\mathbf{b}^f}{\partial X_m} - \bar{\omega}\frac{\partial \mathbf{b}^{f\dagger}\mathbf{S}^{[2]}\mathbf{b}^f}{\partial X_m} - \sum_n \bar{X}_n \frac{\partial (\mathbf{FDS} - \mathbf{SDF})_n}{\partial X_m} = 0, \tag{2.86}$$

where X_m are the orbital rotation parameters. Due to the $2n + 1$ rule, and since the gradient is a first-order property, we only need to solve the above equations through zero order. Eqs. (2.82)-(2.85) are thus already taken care of, and it is seen that the multiplier $\bar{\omega}$ is determined as the eigenvalue of the linear response equations, i.e. it corresponds to the excitation energy. It is then only necessary to determine the Lagrange multipliers $\bar{\mathbf{X}}$ such that Eq. (2.86) is also fulfilled.

2.4.2 The Lagrange Multipliers

To evaluate the terms in Eq. (2.86), the asymmetric Baker-Campbell-Hausdorff (BCH) expansion[46] of the exponentially parameterized density is applied

$$\mathbf{D}(\mathbf{X}) = \exp(-\mathbf{XS})\mathbf{D}\exp(\mathbf{SX}) = \mathbf{D} + [\mathbf{D}, \mathbf{X}]_S + \cdots, \tag{2.87}$$

where

$$[\mathbf{A}, \mathbf{B}]_S = \mathbf{ASB} - \mathbf{BSA}. \tag{2.88}$$

Since the derivatives are evaluated at the expansion point, only terms of first order in $\mathbf{X}$ are non-zero. The last term in Eq. (2.86) is found to be equal to[61]

$$\mathbf{E}^{[2]}\bar{\mathbf{X}} = \mathbf{F}[\bar{\mathbf{X}}, \mathbf{D}]_S\,\mathbf{S} - \mathbf{S}[\bar{\mathbf{X}}, \mathbf{D}]_S\,\mathbf{F} + \mathbf{G}\left([\bar{\mathbf{X}}, \mathbf{D}]_S\right)\mathbf{DS} - \mathbf{SDG}\left([\bar{\mathbf{X}}, \mathbf{D}]_S\right). \tag{2.89}$$

We can thus find $\bar{\mathbf{X}}$ by solving the set of linear equations

$$\mathbf{E}^{[2]}\bar{\mathbf{X}} = \frac{\partial E^0}{\partial \mathbf{X}} + \frac{\partial \mathbf{b}^{f\dagger}\mathbf{E}^{[2]}\mathbf{b}^f}{\partial \mathbf{X}} - \bar{\omega}\frac{\partial \mathbf{b}^{f\dagger}\mathbf{S}^{[2]}\mathbf{b}^f}{\partial \mathbf{X}}. \tag{2.90}$$

From the matrix expressions for $\mathbf{b}^{f\dagger}\mathbf{E}^{[2]}\mathbf{b}^f$ and $\mathbf{b}^{f\dagger}\mathbf{S}^{[2]}\mathbf{b}^f$ [61]

$$\mathbf{b}^{f\dagger}\mathbf{E}^{[2]}\mathbf{b}^{f} = -\mathrm{Tr}\,\mathbf{F}\Big[\Big[\mathbf{b}^{f},\mathbf{D}\Big]_{\mathrm{S}},\mathbf{b}^{f\dagger}\Big]_{\mathrm{S}} - \mathrm{Tr}\,\mathbf{G}\Big(\Big[\mathbf{b}^{f},\mathbf{D}\Big]_{\mathrm{S}}\Big)\Big[\mathbf{D},\mathbf{b}^{f\dagger}\Big]_{\mathrm{S}} \tag{2.91}$$

$$\mathbf{b}^{f\dagger}\mathbf{S}^{[2]}\mathbf{b}^{f} = \mathrm{Tr}\,\mathbf{b}^{f\dagger}\mathbf{S}\Big[\mathbf{D},\mathbf{b}^{f}\Big]_{\mathrm{S}}\mathbf{S} \tag{2.92}$$

and the relations for the two-electron integrals

$$\mathbf{G}^{\mathrm{T}}(\mathbf{A}) = \mathbf{G}\left(\mathbf{A}^{\mathrm{T}}\right) \tag{2.93}$$

$$\mathrm{Tr}\,\mathbf{A}\mathbf{G}(\mathbf{B}) = \mathrm{Tr}\,\mathbf{B}\mathbf{G}(\mathbf{A}), \tag{2.94}$$

the terms on the right hand side of Eq. (2.90) are found as

$$\frac{\partial E^{0}}{\partial \mathbf{X}} = 0, \tag{2.95}$$

$$-\bar{\omega}\frac{\partial \mathbf{b}^{f\dagger}\mathbf{S}^{[2]}\mathbf{b}^{f}}{\partial \mathbf{X}} = -2\bar{\omega}\Big[\mathbf{SDS}\Big[\mathbf{b}^{f},\mathbf{b}^{f\dagger}\Big]_{\mathrm{S}}\mathbf{S}\Big]^{\mathrm{A}}, \tag{2.96}$$

$$\frac{\partial \mathbf{b}^{f\dagger}\mathbf{E}^{[2]}\mathbf{b}^{f}}{\partial \mathbf{X}} = \mathbf{ADS} - \mathbf{SDA}, \tag{2.97}$$

where

$$\begin{aligned}\mathbf{A} = {} & \mathbf{S}\mathbf{b}^{f\dagger}\left(\mathbf{F}\mathbf{b}^{f}\mathbf{S} - \mathbf{S}\mathbf{b}^{f}\mathbf{F}\right) - \left(\mathbf{F}\mathbf{b}^{f}\mathbf{S} - \mathbf{S}\mathbf{b}^{f}\mathbf{F}\right)\mathbf{b}^{f\dagger}\mathbf{S} + \mathbf{G}\left(\Big[\mathbf{b}^{f},\Big[\mathbf{D},\mathbf{b}^{f\dagger}\Big]_{\mathrm{S}}\Big]_{\mathrm{S}}\right) \\ & +2\Big[\mathbf{S}\mathbf{b}^{f\dagger}\mathbf{G}\left(\Big[\mathbf{b}^{f},\mathbf{D}\Big]_{\mathrm{S}}\right) - \mathbf{G}\left(\Big[\mathbf{b}^{f},\mathbf{D}\Big]_{\mathrm{S}}\right)\mathbf{b}^{f\dagger}\mathbf{S}\Big]^{\mathrm{S}}\end{aligned} \tag{2.98}$$

and

$$[\mathbf{M}]^{\mathrm{A}} = \tfrac{1}{2}\mathbf{M} - \tfrac{1}{2}\mathbf{M}^{\dagger} \tag{2.99}$$

$$[\mathbf{M}]^{\mathrm{S}} = \tfrac{1}{2}\mathbf{M} + \tfrac{1}{2}\mathbf{M}^{\dagger}. \tag{2.100}$$

Eq. (2.95) is straight forward since the variational condition Eq. (2.75) is fulfilled at the expansion point.

2.4.3 The Geometrical Gradient

The excited state geometrical gradient should be expressed in terms of the first derivatives of the one and two electron integral matrices $\mathbf{h}^{x}$, $\mathbf{G}^{x}$, $\mathbf{S}^{x}$ and the density, Fock and overlap matrices at the expansion point $\mathbf{x}_0$. The notation $\mathbf{A}^{x}$ denotes the geometrical first derivative of $\mathbf{A}$. In ref. 66 it was found that the first derivative of the density $\mathbf{D}^{x}(\mathbf{X})$ is given by the first derivative of the reference density matrix $\mathbf{D}^{x}$ which, from the idempotency condition for $\mathbf{D}$, is found to be

$$\mathbf{D}^{x} = -\mathbf{D}\mathbf{S}^{x}\mathbf{D}. \tag{2.101}$$

The first-order geometrical derivative is given by

$$\frac{\mathrm{d}E^{f}}{\mathrm{d}x} = \frac{\mathrm{d}L^{f}}{\mathrm{d}x} = \frac{\mathrm{d}E^{0}}{\mathrm{d}x} + \frac{\partial \mathbf{b}^{f\dagger}\mathbf{E}^{[2]}\mathbf{b}^{f}}{\partial x} - \bar{\omega}\frac{\partial \mathbf{b}^{f\dagger}\mathbf{S}^{[2]}\mathbf{b}^{f}}{\partial x} - \bar{\mathbf{X}}\frac{\partial(\mathbf{FDS} - \mathbf{SDF})}{\partial x}. \tag{2.102}$$

The first term is simply the geometrical gradient of the ground state. In ref. 66 this was shown to be

$$E^{0x} = 2\,\mathrm{Tr}\,\mathbf{D}\mathbf{h}^x + \mathrm{Tr}\,\mathbf{D}\mathbf{G}^x(\mathbf{D}) + \mathrm{Tr}\,\mathbf{D}^x\mathbf{F} + h^x_{\mathrm{nuc}}\,. \tag{2.103}$$

The other terms are found as the derivative of the matrix expressions in Eq. (2.91) and (2.92)

$$\begin{aligned}\frac{\partial \mathbf{b}^{f\dagger}\mathbf{E}^{[2]}\mathbf{b}^f}{\partial x} = &-\mathrm{Tr}\left(\mathbf{F}^x + \mathbf{G}\left(\mathbf{D}^x\right)\right)\left[\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}} - \mathrm{Tr}\,\mathbf{F}\left[\left[\mathbf{b}^f,\mathbf{D}^x\right]_{\mathrm{S}},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}}\\ &-\mathrm{Tr}\,\mathbf{F}\left[\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}^x},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}} - \mathrm{Tr}\,\mathbf{F}\left[\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}^x}\\ &-\mathrm{Tr}\,\mathbf{G}^x\left(\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}}\right)\left[\mathbf{D},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}}\\ &-2\,\mathrm{Tr}\,\mathbf{G}\left(\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}}\right)\left(\left[\mathbf{D}^x,\mathbf{b}^{f\dagger}\right]_{\mathrm{S}} + \left[\mathbf{D},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}^x}\right)\end{aligned} \tag{2.104}$$

$$\begin{aligned}-\bar{\omega}\frac{\partial \mathbf{b}^{f\dagger}\mathbf{S}^{[2]}\mathbf{b}^f}{\partial x} = &-\bar{\omega}\,\mathrm{Tr}\,\mathbf{b}^{f\dagger}\mathbf{S}^x\left[\mathbf{D},\mathbf{b}^f\right]_{\mathrm{S}}\mathbf{S}\\ &-\bar{\omega}\,\mathrm{Tr}\,\mathbf{b}^{f\dagger}\mathbf{S}\left(\left[\mathbf{D}^x,\mathbf{b}^f\right]_{\mathrm{S}}\mathbf{S} + \left[\mathbf{D},\mathbf{b}^f\right]_{\mathrm{S}^x}\mathbf{S} + \left[\mathbf{D},\mathbf{b}^f\right]_{\mathrm{S}}\mathbf{S}^x\right)\end{aligned} \tag{2.105}$$

$$-\bar{\mathbf{X}}\frac{\partial\left(\mathbf{FDS}-\mathbf{SDF}\right)}{\partial x} = -2\bar{\mathbf{X}}\left[\mathbf{F}^x\mathbf{DS} + \mathbf{G}\left(\mathbf{D}^x\right)\mathbf{DS} + \mathbf{FD}^x\mathbf{S} + \mathbf{FDS}^x\right]^{\mathrm{A}}, \tag{2.106}$$

where $\mathbf{F}^x = \mathbf{h}^x + \mathbf{G}^x(\mathbf{D})$. Collecting the various terms we obtain

$$\begin{aligned}\frac{\partial E^f}{\partial x} = &\,\mathrm{Tr}\left(2\mathbf{D} - \left[\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}} - \left[\mathbf{D},\bar{\mathbf{X}}\right]_{\mathrm{S}}\right)\mathbf{h}^x - \mathrm{Tr}\left[\mathbf{D},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}}\mathbf{G}^x\left(\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}}\right)\\ &+\mathrm{Tr}\left(\mathbf{D} - \left[\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}} - \left[\mathbf{D},\bar{\mathbf{X}}\right]_{\mathrm{S}}\right)\mathbf{G}^x(\mathbf{D}) + h^x_{\mathrm{nuc}}\\ &-\mathrm{Tr}\,\mathbf{D}^x\mathbf{G}\left(\left[\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}}\right) - 2\,\mathrm{Tr}\left(\left[\mathbf{D}^x,\mathbf{b}^{f\dagger}\right]_{\mathrm{S}} + \left[\mathbf{D},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}^x}\right)\mathbf{G}\left(\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}}\right)\\ &-\mathrm{Tr}\,\mathbf{D}^x\mathbf{G}\left(\left[\mathbf{D},\bar{\mathbf{X}}\right]_{\mathrm{S}}\right) - \mathrm{Tr}\left(\left[\mathbf{D}^x,\bar{\mathbf{X}}\right]_{\mathrm{S}} + \left[\mathbf{D},\bar{\mathbf{X}}\right]_{\mathrm{S}^x}\right)\mathbf{F}\\ &-\mathrm{Tr}\left(\left[\left[\mathbf{b}^f,\mathbf{D}^x\right]_{\mathrm{S}},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}} + \left[\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}^x},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}} + \left[\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}^x}\right)\mathbf{F}\\ &+\omega_f\,\mathrm{Tr}\,\mathbf{b}^{f\dagger}\mathbf{S}\left(\left[\mathbf{b}^f,\mathbf{D}^x\right]_{\mathrm{S}}\mathbf{S} + \left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}^x}\mathbf{S} + \left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}}\mathbf{S}^x\right)\\ &+\omega_f\,\mathrm{Tr}\,\mathbf{b}^{f\dagger}\mathbf{S}^x\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}}\mathbf{S},\end{aligned} \tag{2.107}$$

where $\mathbf{G}\left(\left[\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}},\mathbf{b}^{f\dagger}\right]_{\mathrm{S}}\right)$, $\mathbf{G}\left(\left[\mathbf{D},\bar{\mathbf{X}}\right]_{\mathrm{S}}\right)$, $\mathbf{G}\left(\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}}\right)$ and $\mathbf{F}$ can be evaluated, whereas $\mathbf{G}^x(\mathbf{D})$, $\mathbf{G}^x\left(\left[\mathbf{b}^f,\mathbf{D}\right]_{\mathrm{S}}\right)$, $\mathbf{h}^x$ and h^x_{nuc} have to be evaluated for each geometrical perturbation.

Note that no two-electron integrals are represented explicitly, in order to obtain the best performance – e.g. for linear scaling codes - no reference should be made to four-index integrals.

2.4.4 The First-order Excited State Properties

The expression for the first-order one-electron excited state properties for perturbation independent basis sets is obtained from the expression for the excited state gradient by omitting all two-electron derivative terms, as well as all terms involving the derivative of the overlap matrix

$$\left\langle f \left| h^x \right| f \right\rangle = 2\,\mathrm{Tr}\,\mathbf{D}\mathbf{h}^x - \mathrm{Tr}\left(\left[\left[\mathbf{b}^f, \mathbf{D} \right]_S, \mathbf{b}^{f\dagger} \right]_S - \left[\mathbf{D}, \bar{\mathbf{X}} \right]_S \right) \mathbf{h}^x + h_{nuc}^x \,. \tag{2.108}$$

The first and last terms in Eq. (2.108) correspond to the ground state first order property as seen from Eq. (2.103).

2.5 Test Calculations

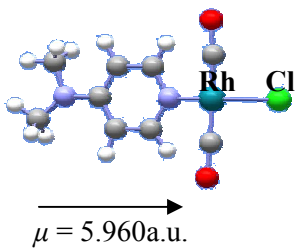
To illustrate the possibilities of an AO response solver in connection with our SCF optimization program, test calculations have been carried out on problematic cases from the first part of the thesis. The lowest excitation energy and the average polarizability, both static and in a field with ω = 0.03a.u., have been found for the zinc complex in Fig. 1.3 and the rhodium complex in Fig. 1.33. The levels of theory chosen are those where DIIS could not optimize the reference state, namely LDA/6-31G for the zinc complex and HF/AhlrichsVDZ with STO-3G on the rhodium for the rhodium complex.

Table 2-1 Ground state properties obtained with our AO response solver. All numbers are in a.u.

		The average polarizability		Excitation energy
		static	ω = 0.03	
Rhodium complex	HF/AhrichsVDZ	170.598	173.349	0.0938
Zinc complex	LDA/6-31G	161.406	162.517	0.0713

The basis sets applied in the test calculations are not satisfactory for serious polarizability calculations, and the numbers only demonstrate the perspectives of the AO response solver in combination with the SCF optimization algorithms described in Part 1. When the solver is fully implemented in the AO basis, we will be able to obtain molecular properties for large complex molecules in a routine manner.

The implementation of the excited state gradient is a work in progress. So far we have implemented calculation of first-order one-electron properties of the excited state for perturbation independent basis sets as described in Section 2.4.4. The excited state dipole moment of the Rhodium complex from above has been found as



Again it should be noted that the basis set is insufficient for this type of calculation. This is only to demonstrate that it can be done.

2.6 Conclusion

The atomic orbital (AO) based response equations have been derived using the second quantization framework. In particular, the proof of pairing is considered. Since the diagonal elements in $\boldsymbol{\kappa}$ are not redundant in the AO basis, the proof given in the MO basis cannot be directly applied. However, it is shown that there is also pairing in the AO basis.

An AO response solver has been implemented similar to the solver in the MO basis with a few exceptions. The lack of diagonal dominance in the electronic Hessian in the AO basis makes preconditioning a difficult task. Optimally, the AO solver should be implemented in a linear scaling manner with only matrix multiplications and additions, and without reference to the MO basis. However, currently a transformation is made to the MO basis where the preconditioning is carried out followed by a transformation back to the AO basis. The redundant orbital rotations, which are simply left out of the MO equations, are removed in the AO formulation using projection operators.

The response equations and molecular property expressions are simpler in the AO formulation than in the MO formulation. To demonstrate how expressions for properties can easily be derived in the AO response framework, the expression for the geometrical gradient of the singlet excited state has been derived.

To illustrate the possibilities of the AO optimization methods presented in Part 1, joined with the AO response solver presented in this part of the thesis, test calculations are given for cases where DIIS diverged when optimizing the reference state. The averaged polarizability and the lowest excitation energy are given as well as the excited state dipole for one of the examples.

The derivation and implementation of the various molecular properties is straightforward in the AO formulation compared to the MO formulation as exemplified by the excited state geometrical gradient. Especially the derivation of higher derivatives of molecular properties is simplified, and it will thus be natural to expand our response program in this direction. However, before calculations of molecular properties of large and complex molecules can be carried out in a truly linear scaling framework, the problems related to preconditioning of the AO solver must be solved.

Part 3

Benchmarking for Radicals

3.1 Introduction

To corroborate the reliability of *ab initio* quantum chemical predictions of molecular properties, it is important to investigate and describe strengths and weaknesses of the many-electron models through systematic benchmark studies on different kinds of molecules.

Regarding open-shell molecules, benchmarks have been reported comparing open- and closed-shell molecules examining the accuracy of molecular properties computed by various many-electron models. In a study of the atomization energies of 11 small molecules[67] no significant difference in the performance for closed- and open-shell molecules was found for the CCSDT model. However, in another study[68] it was found that even though the CCSD(T) model performs convincingly for closed-shell molecules, the performance for open-shell molecules is less impressive.

In this part of the thesis full configuration interaction (FCI) benchmarks of molecular properties for the small open-shell molecules CN and CCH are presented. In the FCI model, all Slater determinants arising from distributing the electrons in the given one-electron basis with correct symmetry and spin-projection are included. Errors due to truncation of the many-electron basis are thus eliminated in an FCI calculation and it provides important benchmarks for other many-electron models. For open-shell molecules, the number of FCI benchmarks is limited and the work presented in this part of the thesis is an attempt to improve on this situation. We thus hope our results will serve as valuable benchmarks for further analysis of open-shell methods.

3.2 Computational Methods

All calculations have been carried out with the quantum chemical program package LUCIA[69], using integrals and Hartree-Fock (HF) orbitals obtained from the DALTON[70] program. The calculations

are based on a ROHF reference wave function, but no spin-adaption is imposed in the CI and CC calculations.

All FCI calculations have been carried out in the Dunnings cc-pVDZ[71] basis set. Since the number of determinants in the FCI model increases exponentially with the number of basis functions and electrons, it is currently not feasible to do the FCI calculations on CN and CCH in the cc-pVTZ basis. As the cc-pVDZ basis does not provide accurate geometries and energetics,[46] we will also obtain the equilibrium geometry, harmonic frequency, and dissociation energy for CN using the cc-pVTZ[71] basis set in coupled cluster calculations, including up to quadruple excitations. In addition, FCI and CC calculations up to quadruples level have been carried out on CN and CN^- in the basis set aug-cc-pVDZ without the diffuse d-functions (aug´-cc-pVDZ) to obtain the vertical electron affinity of CN.

We investigate two ways of defining the excitation-level in CC. The typical approach is to let the excitation level identify the allowed number of orbital excitations, denoted CC(orb). If instead the excitation level is taken to identify the spin-orbital excitation level, selected excitations, which involve spin-flipping and other internal excitations, are excluded from the calculation for open-shell molecules. This scheme will be referred to as CC(spin-orb). The difference between the two definitions of the excitation level is illustrated in Fig. 3.1. The CI calculations will all be carried out with orbital excitations.

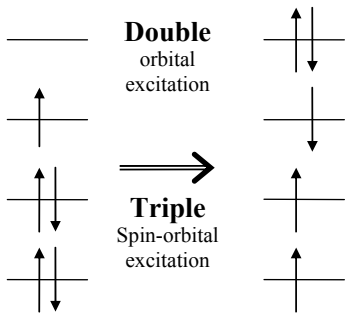


Fig. 3.1 An excitation which would be included in a CCSD(orb) calculation, but not in a CCSD(spin-orb) calculation.

In the following SD, SDT, SDTQ, SDTQ5, SDTQ56 and SDTQ567 denote excitation-spaces which include up to 2, 3, 4, 5, 6 and 7 excitations from the occupied spin-orbitals respectively.

3.3 Numerical Results

First, the convergence of the CC and CI hierarchies for the open shell molecule CN is studied. Next, the potential curve for CN is obtained from CCSD, CCSDT, CCSDTQ, and FCI calculations at various inter-nuclear distances. In Section 3.3.3, the equilibrium geometries, harmonic frequencies, and dissociation energies obtained for CN are presented and in Section 3.3.4 the vertical electron affinity for CN is found. Finally, in Section 3.3.5 a minor benchmark study is presented where the equilibrium geometry of the intergalactic radical CCH is determined at the FCI level.

3.3.1 Convergence of CC and CI Hierarchies

The convergence of the CC and CI hierarchies are studied. For CN calculations have been carried out at the experimental equilibrium distance[72] r_{exp} = 1.1718Å at the levels CCSD through CCSDTQ56. Both the orbital excitation and spin-orbital excitation approaches are considered. In addition, calculations have been carried out at the levels CISD through CISDTQ567 and in FCI. In all calculations the cc-pVDZ basis-set is used. The results are seen in Fig. 3.2.

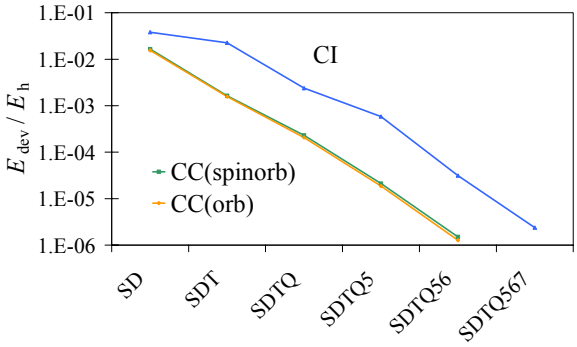


Fig. 3.2 E_{dev} for CC with spin-orbital and orbital excitation levels and for CI with orbital excitation levels. $E_{dev} = E - E_{FCI}$.

The first thing to note is the similarity of the two CC curves. Clearly the spin-orbital excitation restriction does not affect the accuracy in a significant way, the deviation energies are in all cases smaller for CC(orb), but the difference is negligible.

Comparing the CI curve with the CC curves, two trends are obvious; the smooth convergence of the CC hierarchy compared to the CI hierarchy and the faster convergence of the CC hierarchy. The CC energy obtained using up to *n*-fold excitations is roughly as accurate as the CI energy using up to *n*+1-fold excitations. Both phenomena are explained by the inclusion of disconnected clusters in the CC wave function. At a given level of CC theory, the CC wave function includes all the CI configurations at the same level of CI theory plus some higher excitations arising from disconnected clusters. Consequently, it covers the dynamical correlation better than CI and is thus at the given

level closer to the FCI solution. Describing the convergence pattern of the CI and CC hierarchies through orders of Møller-Plesset perturbation theory (MPPT),[73] the form of the curves can be predicted. Because also disconnected products of excitations are included in the ansatz of CC, the order of its error grows continually in the order of MPPT. Going from uneven to even excitation levels, both methods have an increase in the order of error in energy of two orders of MPPT, thus, the graphs are parallel. Going from even to uneven excitation levels, the CC error increases one order, whereas the CI error remains unchanged, giving a greater slope for the CC curve. This explains the parallel behavior going from uneven to even excitation levels and the smoother convergence of the CC hierarchy compared to the CI hierarchy. The stepwise convergence predicted by MPPT, which should be significant for CI and noticeably for CC, is not apparent though. The reason could be that CN is not strictly mono-configurational.

The convergence patterns for CI and CC are very similar to the convergence patterns previously reported for N_2.[74] Therefore, it does not seem that the open-shell nature of CN leads to slow convergence of the CI and CC hierarchies compared to closed shell cases.

3.3.2 The Potential Curve for CN

The potential curve for CN was determined from single-point calculations at the FCI level with basis set cc-pVDZ. Close to equilibrium the energies were converged to 10^{-9} E_h making the determination of accurate spectroscopic constants possible. The result is displayed in Fig. 3.3.

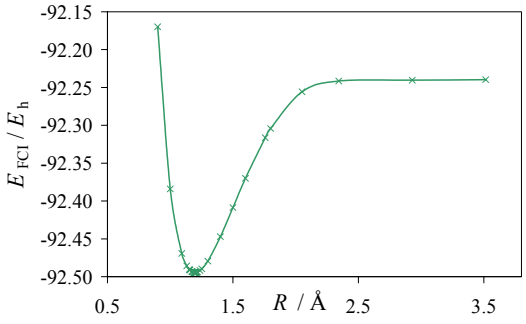


Fig. 3.3 The potential curve for CN found from FCI cc-pVDZ calculations.

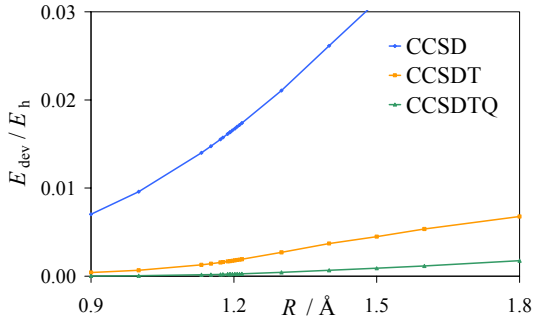


Fig. 3.4 E_{dev} for the CC potential curves. $E_{dev}(R) = E(R) - E_{FCI}(R)$.

The potential curve was also created with the methods CCSD(orb), CCSDT(orb) and CCSDTQ(orb) in the basis set cc-pVDZ. Since the weight of the reference HF- determinant decreases as the inter-nuclear distance increases, we examine the HF-coefficients from the FCI calculations and discover that it is irrelevant to make single-reference CC calculations beyond R = 1.8Å, since the weight of the reference has already dropped to 0.57 at that point. Fig. 3.4 displays the differences of the CC

potential curves compared to the FCI curve. At a given inter-nuclear distance, the FCI energy has been subtracted from the CC energy.

The decreasing weight of the reference ground state with increasing atomic distance is reflected in the quality of the CC wave functions. The correlation in the wave function compensates partially for the lack of a single dominant configuration; the higher the correlation level, the better the compensation. This is illustrated by the slopes of the curves in Fig. 3.4. Furthermore, it should be noticed how the deviation energy is nearly linear in R, with a slightly positive curvature around the equilibrium geometry.

3.3.3 Spectroscopic Constants and Atomization Energy for CN

The equilibrium geometry and harmonic frequency for CN were found from single-point calculations using quartic interpolation. The atomization energy was found at the experimental equilibrium distance. The results are displayed in Table 3-1.

Table 3-1 Equilibrium geometry, harmonic frequency, and atomization energy for CN.

		R_{eq} / Å	ω_e / cm^{-1}	D_e / kJ/mol
CCSD(spin-orb)	cc-pVDZ	1.1855	2114	629.2
CCSD(orb)	cc-pVDZ	1.1860	2111	631.6
CCSDT(spin-orb)	cc-pVDZ	1.1944	2046	662.9
CCSDT(orb)	cc-pVDZ	1.1946	2043	663.0
CCSDTQ(spin-orb)	cc-pVDZ	1.1964	2026	666.4
CCSDTQ(orb)	cc-pVDZ	1.1964	2025	666.5
FCI	cc-pVDZ	1.1969	2020	667.0
CCSD(spin-orb)	cc-pVTZ	1.1688	2136	674.2
CCSDT(spin-orb)	cc-pVTZ	1.1783	2067	714.4
CCSDTQ(spin-orb)	cc-pVTZ	1.1804	2045	718.5
Experimental[72]		1.1718	2069	---

As mentioned in Section 3.2, it is not feasible to carry out FCI calculations at the cc-pVTZ level. Still, the convergence of the CC hierarchy can be estimated by examining the changes in the constants. Since the difference in accuracy between the models CC(orb) and CC(spin-orb) is negligible compared to the deviation from FCI, only the CC(spin-orb) results are discussed from now on and only the CC(spin-orb) numbers are found at the cc-pVTZ level.

The deviation curves for the coupled cluster energies (see Fig. 3.4) are increasing functions, and thus the coupled cluster equilibrium bond lengths are shorter than the one found from FCI. Furthermore, the positive curvature of the deviation-curves around the equilibrium leads to coupled cluster frequencies that are higher than the FCI frequency.

As expected, the cc-pVDZ basis set does not provide accurate geometries and frequencies, and the cc-pVTZ numbers are clearly more in the range of the experimental data than the cc-pVDZ numbers.

CCSD displays its insufficiency for prediction of equilibrium properties by differing from the FCI values by 0.01Å in the geometry, 90 cm^{-1} in the frequency, and 35 kJ/mol in the atomization energy. The errors in R_{eq} and ω_e are reduced by a factor of four going to the CCSDT level and a factor of five going from the CCSDT to the CCSDTQ level. The error in the atomization energy is reduced by a factor of nine going to the CCSDT level and a factor of eight going from the CCSDT to the CCSDTQ level, but while the equilibrium geometry on the CCSDTQ level is only 0.0005Å from the FCI value, the harmonic frequency is still about 5 cm^{-1} too high.

Both the equilibrium geometry and the harmonic frequency are apparently better approximated by the CCSDT method than the CCSDTQ. This is due to a favorable cancellation in errors for CCSDT calculations in small basis sets. By extrapolation to the larger aug-cc-pVQZ basis,[67,75] we get an equilibrium distance of 1.1759Å and a harmonic frequency of 2060cm^{-1} at the CCSDTQ level.

3.3.4 The Vertical Electron Affinity of CN

Calculations on CN^- and CN were carried out in the aug´-cc-pVDZ basis at the experimental equilibrium geometry for CN. The FCI calculation on CN^- is one of the largest FCI calculations carried out so far containing about 20 billion Slater determinants. The vertical electron affinity (EA) was found and is displayed in Table 3-2. Again only CC(spin-orb) calculations have been carried out because of the rather small difference in performance of CC(spin-orb) and CC(orb).

Table 3-2 The vertical electron affinity of CN.

		EA / E_h	EA - EA_{FCI}
CCSD(spin-orb)	aug'-cc-pVDZ	0.13025	0.00063
CCSDT(spin-orb)	aug'-cc-pVDZ	0.12977	0.00014
CCSDTQ(spin-orb)	aug'-cc-pVDZ	0.12966	0.00003
FCI	aug'-cc-pVDZ	0.12962	---

The convergence is remarkable; already at the CCSD level we are down to an error of 0.5% of the FCI value, on the CCSDT level it is 0.1% and on the CCSDTQ level 0.02%. The reason for the excellent convergence is found in a cancellation of errors that influence the result. The deviations of the individual energies are always roughly an order of magnitude larger than the deviation of the affinity,[75] but the errors cancel when the CN and CN^- energies are subtracted. That the convergence is from above is also noteworthy. This is because the CC hierarchy converges faster for CN^- than for

CN. This seems surprising since CN^- contains one more electron than CN, but it could be explained by CN^- being more one-configurational than CN.

3.3.5 The Equilibrium Geometry of CCH

The equilibrium geometry of CCH found from FCI/cc-pVDZ calculations is used in ref. 76 to calibrate coupled cluster calculations in larger basis sets. The FCI correction is assumed to be independent of basis set.

To optimize for the two variables R(CC) and R(CH), the CCH radical is assumed linear and the CC and CH bonds are then distorted in step-lengths of $\delta = 0.01$Å from an initial geometry making a grid of single-point calculations around the equilibrium geometry with R(CC) on the one axis and R(CH) on the other. The initial geometry is taken from a CCSDT cc-pVDZ study[76], the geometry being R^{CCSDT}(CC) = 1.23448Å and R^{CCSDT}(CH) = 1.07924Å. The resulting potential energy surface is seen in Fig. 3.5.

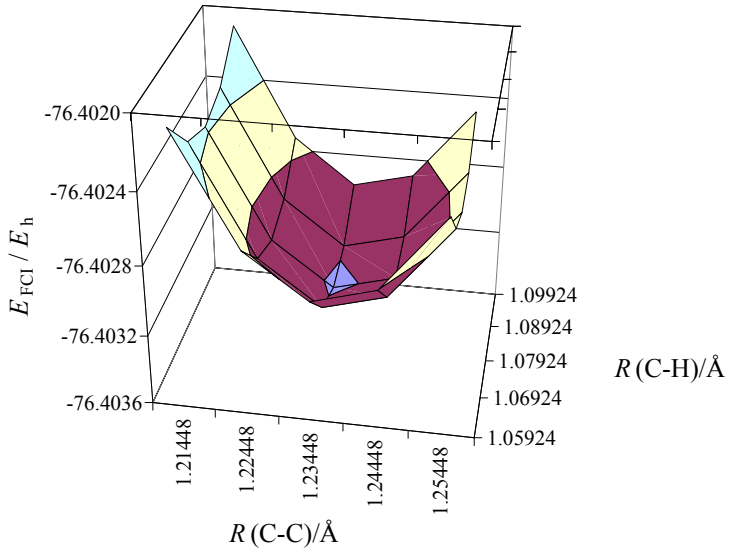


Fig. 3.5 The potential energy surface of CCH.

From finite-difference expressions with the error being of the order δ^4, the gradient and Hessian are found for the initial geometry and a Newton step is taken giving an improved guess for the equilibrium geometry. The FCI equilibrium geometry is thus found as

$$\mathbf{R}_{FCI} = \mathbf{R}^{CCSDT} - \mathbf{H}^{-1}\mathbf{G}\,, \tag{3.1}$$

where $\mathbf{G}$ is the gradient, $\mathbf{H}$ the Hessian, and $\mathbf{R}^{CCSDT}$ the CCSDT geometry.

The equilibrium geometry at the FCI level is found to be

$$R_{\mathrm{FCI}}(\mathrm{CC}) = 1.2367\text{Å} \text{ and } R_{\mathrm{FCI}}(\mathrm{CH}) = 1.0802\text{Å}.$$

The error in the resulting geometry is a sum of the error from the finite difference approximations and the error from the Newton step. The gradient and Hessian carry an error of $\mathcal{O}(\delta^4)$ where δ = 0.01Å, this is an error in the order of 10^{-8}Å. The Newton step has an error of $\mathcal{O}((\mathbf{H}^{-1}\mathbf{G})^2)$, in this case $\mathbf{H}^{-1}\mathbf{G}$ is of the size 10^{-3}Å and so the error is in the order of 10^{-6}Å. The error in total is thus in the order of 10^{-6}Å.

The gradient for the FCI equilibrium geometry has been found as above, making single-point calculations at the FCI geometry and at geometries distorted in steps of 0.01Å from the FCI geometry. The same finite-difference expressions as before are used. The gradient is found to be

$$\mathbf{G}_{\mathrm{FCI}} = \left[-1.8593 \cdot 10^{-5} \, E_{\mathrm{h}} \middle/ \text{Å} ; 3.0661 \cdot 10^{-5} \, E_{\mathrm{h}} \middle/ \text{Å} \right], \tag{3.2}$$

thus verifying the correctness of the FCI geometry.

Since the geometry was determined at the CCSDT level to be $R^{\mathrm{CCSDT}}(\mathrm{CC})$ = 1.23448Å and $R^{\mathrm{CCSDT}}(\mathrm{CH})$ = 1.07924Å, the error due to truncation of the many-electron basis in CCSDT is in the order of 10^{-3}Å. This is similar to the results obtained for CN. This also suggests that the quadruples correction to the equilibrium geometry is in the order of 0.001-0.002Å.

3.4 Conclusion

Full configuration interaction (FCI) and coupled cluster (CC) calculations have been carried out on CN using the cc-pVDZ and cc-pVTZ basis sets. The equilibrium bond distance, harmonic frequency, atomization energy, and vertical electron affinity have been evaluated on the various levels of theory.

As expected, the cc-pVDZ basis set does not provide accurate geometries and frequencies and CCSD is insufficient for prediction of equilibrium properties. Apparently, the CCSDT method is a better approximation than CCSDTQ for obtaining the equilibrium geometry and the harmonic frequency. This is due to a favorable cancellation of errors for CCSDT calculations in small basis sets. Also the vertical electron affinities are affected by cancellation of errors, and already at the CCSD level, the error is less than 1mE_{h} compared to the FCI value.

The convergence patterns for the CI and CC hierarchies are studied for CN and it is found similar to the convergence patterns previously reported for N_2.[74] Thus, it does not seem that the open-shell nature of CN leads to slow convergence of the CI and CC hierarchies compared to closed shell cases.

For a number of the CC calculations, the excitation levels have been defined by spin-orbital excitations instead of orbital excitations. Certain internal excitations are thereby omitted, but it is seen that this does not affect the accuracy in any significant way. For a given excitation level, the energies obtained in the orbital formalism are in all cases closer to the FCI energy than the ones obtained in the spin-orbital formalism. However, the difference is negligible.

The equilibrium geometry of CCH has been found at the FCI level in the cc-pVDZ basis set to be $R_{FCI}(CC) = 1.2367$Å and $R_{FCI}(CH) = 1.0802$Å. The correction found to the initial CCSDT geometry is in the order of 10^{-3}Å. The FCI correction to the CCSDT equilibrium geometry of CN was of the same order.

Summary

The developments in computer hardware and linear scaling algorithms over the last decade have made it possible to carry out *ab-initio* quantum chemical calculations on bio-molecules with hundreds of amino acids and on large molecules relevant for nano-science. Quantum chemical calculations are thus evolving to become a widespread tool for use in several scientific branches. It is therefore important that the algorithms work as black-boxes, such that the user outside quantum chemistry does not have to be concerned with the details of the calculations. In particular Hartree Fock (HF) and density functional theory (DFT) methods are employed for calculations on large systems as they represent good compromises between relatively low computational costs and reasonable accuracy of the results. The HF and DFT methods have been a fundamental part of quantum chemistry for many years, and calculations on molecules of ever increasing size and complexity are made possible due to increasing computer resources. The conventional algorithms used for optimization of the one-electron density in HF and DFT are therefore continually tried on their stability and general performance and occasionally they break down. In these cases the calculation takes more time to complete than acceptable or no result can be obtained at all.

We have improved on this situation. In the first part of this thesis, algorithms are presented which improve the optimization in HF and DFT significantly. The optimization has become more effective and where the optimization broke down using conventional algorithms, it now converges without problems. Furthermore, the presented algorithms have no problem-specific parameters and can thus be used as black-boxes.

When the one-electron density has been optimized, molecular properties such as polarizabilities and excitation energies can be calculated. Response theory is often used for this purpose. In the second part of this thesis an atomic orbital (AO) based formulation of response theory is presented which allows linear scaling calculations of molecular properties. Furthermore, the derivation of expressions for molecular properties is simpler in the AO formulation than in the molecular orbital formulation typically used. To illustrate the benefits, the expression for the geometrical derivative of the excited state is derived in the AO formulation.

To confirm the reliability of quantum chemical predictions of molecular properties, it is important to investigate and describe strengths and weaknesses of the quantum chemical models employed. The full configuration interaction (FCI) model is exact within a certain basis set of atomic orbitals. It is thus of great value to be able to compare results from approximate models with FCI results. In the third part of this thesis FCI results are presented for two open-shell molecules, namely CN and CCH. The FCI results are compared with results from approximate models used today for calculations where an accuracy comparable to the experimental is needed.

Dansk Resumé

Udviklingen i det seneste årti indenfor computerhardware og lineært skalerende algoritmer har gjort det muligt at udføre *ab-initio* kvantekemiske beregninger på bio-molekyler med hundredvis af aminosyrer og på store molekyler relevant for nanoteknologi. Kvantekemiske beregninger udvikler sig derfor til at være et bredt anvendt værktøj til brug for adskillige naturvidenskabelige grene. Det er derfor vigtigt at algoritmerne fungerer som såkaldte black-boxes, således at brugere uden for kvantekemi ikke behøver bekymre sig om detaljerne i beregningen. Især Hartree Fock (HF) og density functional theory (DFT) metoderne er benyttet til beregninger på store systemer, da de repræsenterer et godt kompromis mellem fornuftig nøjagtighed af resultaterne og relativ kort beregningstid. HF og DFT er metoder, som har været anvendt i kvantekemien igennem mange år, og da stadig større computer ressourcer er til rådighed bliver de brugt til at udføre beregninger på stadigt større og mere komplekse molekyler. De algoritmer som benyttes i dag til optimering af den en-elektroniske densitet i HF og DFT bliver derfor til stadighed testet på deres stabilitet og effektivitet og til tider bryder de sammen. I disse tilfælde tager beregningen enten uacceptabelt lang tid eller opgiver at levere et resultat.

Vi har forbedret denne situation. I den første del af afhandlingen præsenteres algoritmer, som signifikant forbedrer optimeringen i HF og DFT. Optimeringen er blevet mere effektiv, og tilfælde hvor optimeringen før brød sammen kan nu udføres uproblematisk. De præsenterede algoritmer har desuden ingen problem-specifikke parametre og kan derfor betragtes som black-boxes.

Når den en-elektroniske densitet er optimeret, kan molekylære egenskaber såsom polarisabiliteter og eksitationsenergier beregnes. Til det formål benyttes ofte responsteori. I anden del af afhandlingen præsenteres en atomorbitalformulering af responsteori, som muliggør en lineær skalering af egenskabsberegningerne. Desuden er udviklingen af udtryk for molekylære egenskaber blevet simplere i atomorbitalformuleringen sammenlignet med molekylorbitalformuleringen som ellers typisk benyttes. For at illustrere fordelene er udtrykket for den eksiterede tilstands geometriske gradient udviklet i atomorbitalformuleringen.

For at bekræfte troværdigheden af kvantekemiske forudsigelser af molekylære egenskaber, er det vigtigt at undersøge og beskrive styrker og svagheder ved de kvantekemiske modeller som anvendes. Full configuration interaction (FCI) er en eksakt model inden for et bestemt sæt af atomorbital basisfunktioner. Det er derfor værdifuldt at kunne sammenligne resultater fra approksimative modeller med FCI resultater. I tredje del af afhandlingen er FCI resultater præsenteret for to åben-skal molekyler, CN og CCH. Disse resultater er sammenlignet med resultater fra approksimative modeller, som i dag bruges til at levere kvantekemiske beregninger med en nøjagtighed, som i visse tilfælde overgår den eksperimentelle.

Appendix A

The Derivatives of the DSM Energy

The first and second derivatives of the DSM energy model with respect to **c** is found recalling that

$$E^{\mathrm{DSM}}(\mathbf{c}) = E\left(\bar{\mathbf{D}}\right) + 2\,\mathrm{Tr}\,\bar{\mathbf{F}}\mathbf{D}_{\delta}\,, \tag{A-1}$$

$$E\left(\bar{\mathbf{D}}\right) = E\left(\mathbf{D}_0\right) + 2\,\mathrm{Tr}\,\mathbf{D}_{+}\mathbf{F}_0 + \mathrm{Tr}\,\mathbf{D}_{+}\mathbf{F}_{+}\,, \tag{A-2}$$

$$\mathbf{D}_{+} = \sum_{i=1}^{n} c_i\left(\mathbf{D}_i - \mathbf{D}_0\right), \tag{A-3}$$

and

$$D_{\delta} = 3\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}} - 2\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}} - \bar{\mathbf{D}}\,. \tag{A-4}$$

The two terms in Eq. (A-1) is evaluated one by one:

$$\frac{\partial E\left(\bar{\mathbf{D}}\right)}{\partial c_x} = \mathrm{Tr}\,\mathbf{D}_x\mathbf{F}_0 - \mathrm{Tr}\,\mathbf{D}_0\mathbf{F}_x + \mathrm{Tr}\,\bar{\mathbf{D}}\mathbf{F}_x + \mathrm{Tr}\,\mathbf{D}_x\bar{\mathbf{F}} - \mathrm{Tr}\,\bar{\mathbf{D}}\mathbf{F}_0 - \mathrm{Tr}\,\mathbf{D}_0\bar{\mathbf{F}} \tag{A-5}$$

and

$$\begin{aligned}\frac{\partial}{\partial c_x}2\,\mathrm{Tr}\,\bar{\mathbf{F}}\mathbf{D}_{\delta} &= 2\,\mathrm{Tr}\,\frac{\partial\bar{\mathbf{F}}}{\partial c_x}\mathbf{D}_{\delta} + 2\,\mathrm{Tr}\,\bar{\mathbf{F}}\frac{\partial\mathbf{D}_{\delta}}{\partial c_x}\\ &= 2\,\mathrm{Tr}\,\mathbf{F}_x\mathbf{D}_{\delta} + 2\,\mathrm{Tr}\,\bar{\mathbf{F}}\frac{\partial\mathbf{D}_{\delta}}{\partial c_x}\,,\end{aligned} \tag{A-6}$$

where

$$\frac{\partial\mathbf{D}_{\delta}}{\partial c_x} = 3\bar{\mathbf{D}}\mathbf{S}\mathbf{D}_x + 3\mathbf{D}_x\mathbf{S}\bar{\mathbf{D}} - 2\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}}\mathbf{S}\mathbf{D}_x - 2\bar{\mathbf{D}}\mathbf{S}\mathbf{D}_x\mathbf{S}\bar{\mathbf{D}} - 2\mathbf{D}_x\mathbf{S}\bar{\mathbf{D}}\mathbf{S}\bar{\mathbf{D}} - \mathbf{D}_x\,. \tag{A-7}$$

The second derivative is found in the same manner

$$\frac{\partial^2 E\left(\bar{\mathbf{D}}\right)}{\partial c_x\partial c_y} = 2\,\mathrm{Tr}\,\mathbf{D}_0\mathbf{F}_0 + \mathrm{Tr}\,\mathbf{D}_x\mathbf{F}_y + \mathrm{Tr}\,\mathbf{D}_y\mathbf{F}_x - \mathrm{Tr}\,\mathbf{D}_0\mathbf{F}_x - \mathrm{Tr}\,\mathbf{D}_x\mathbf{F}_0 - \mathrm{Tr}\,\mathbf{D}_y\mathbf{F}_0 - \mathrm{Tr}\,\mathbf{D}_0\mathbf{F}_y\,, \tag{A-8}$$

$$\frac{\partial^2}{\partial c_x\partial c_y}2\,\mathrm{Tr}\,\bar{\mathbf{F}}\mathbf{D}_{\delta} = 2\,\mathrm{Tr}\,\mathbf{F}_x\frac{\partial\mathbf{D}_{\delta}}{\partial c_y} + 2\,\mathrm{Tr}\,\mathbf{F}_y\frac{\partial\mathbf{D}_{\delta}}{\partial c_x} + 2\,\mathrm{Tr}\,\bar{\mathbf{F}}\frac{\partial^2\mathbf{D}_{\delta}}{\partial c_x\partial c_y}\,, \tag{A-9}$$

where

$$\begin{aligned}\frac{\partial^2\mathbf{D}_{\delta}}{\partial c_x\partial c_y} &= 3\mathbf{D}_y\mathbf{S}\mathbf{D}_x + 3\mathbf{D}_x\mathbf{S}\mathbf{D}_y - 2\bar{\mathbf{D}}\mathbf{S}\mathbf{D}_y\mathbf{S}\mathbf{D}_x - 2\mathbf{D}_y\mathbf{S}\bar{\mathbf{D}}\mathbf{S}\mathbf{D}_x - 2\bar{\mathbf{D}}\mathbf{S}\mathbf{D}_x\mathbf{S}\mathbf{D}_y\\ &\quad -2\mathbf{D}_y\mathbf{S}\mathbf{D}_x\mathbf{S}\bar{\mathbf{D}} - 2\mathbf{D}_x\mathbf{S}\bar{\mathbf{D}}\mathbf{S}\mathbf{D}_y - 2\mathbf{D}_x\mathbf{S}\mathbf{D}_y\mathbf{S}\bar{\mathbf{D}}\,.\end{aligned} \tag{A-10}$$

Appendix B

The Density Matrix in the Atomic Orbital Basis

In this appendix we will briefly review the density matrix in the atomic orbital basis and derive the most important relations. For convenience consider a single-determinant wave function with n molecular orbitals occupied. The expectation value of a one-electron operator may then be written as a sum over occupied spin-orbitals

$$\left\langle 0\left|\hat{h}\right|0\right\rangle = \sum_{i=1}^{n} h_{ii} \ . \tag{B-1}$$

Explicitly introducing the MO-AO transformation matrix **C** allow us to write the expectation value as

$$\begin{aligned}\left\langle 0\left|\hat{h}\right|0\right\rangle &= \sum_{i=1}^{n} h_{ii} \\ &= \sum_{\mu,\nu=1}^{N} h_{\mu\nu}\left(\sum_{i=1}^{n} C_{\mu i}^{*} C_{\nu i}\right) \\ &= \sum_{\mu,\nu=1}^{N} h_{\mu\nu} D_{\mu\nu} \quad ,\end{aligned} \tag{B-2}$$

where N is the number of AO basis functions and we have introduced **D** as

$$D_{\mu\nu} = \sum_{i=1}^{n} C_{\mu i}^{*} C_{\nu i} \ . \tag{B-3}$$

It is of interest to study the relation between **D** and the expectation values Δ of Eq. (2.10). To accomplish this we consider the second quantization expression for $\left\langle 0\left|\hat{h}\right|0\right\rangle$ in the nonorthogonal atomic orbital basis. According to ref. 46 one obtains

$$\begin{aligned}\left\langle 0\left|\hat{h}\right|0\right\rangle &= \sum_{\mu,\nu=1}^{N} \left(\mathbf{S}^{-1}\mathbf{h}\mathbf{S}^{-1}\right)_{\mu\nu} \left\langle 0\left|a_{\mu}^{\dagger} a_{\nu}\right|0\right\rangle \\ &= \sum_{\mu,\nu=1}^{N} \left(\mathbf{S}^{-1}\mathbf{h}\mathbf{S}^{-1}\right)_{\mu\nu} \Delta_{\mu\nu} \\ &= \sum_{\mu,\nu=1}^{N} h_{\mu\nu} \left(\mathbf{S}^{-1}\mathbf{\Delta}\mathbf{S}^{-1}\right)_{\mu\nu} \ .\end{aligned} \tag{B-4}$$

By comparing Eqs. (B-4) and (B-2) we have the identification

$$\mathbf{D} = \mathbf{S}^{-1}\mathbf{\Delta}\mathbf{S}^{-1} \ . \tag{B-5}$$

Thus, the density element $\mathbf{D}_{\mu\nu}$ is only identical to the matrix element $\mathbf{\Delta}_{\mu\nu}$ in an orthonormal basis. Although it could be argued that it would be appropriate to call $\mathbf{\Delta}$ the one-electron density matrix in the AO-basis, we will be consistent with the standard literature and call $\mathbf{D}$ the density matrix in the AO basis, and $\mathbf{\Delta}$ the matrix of expectation values of creation-annihilation operators. From the properties of the one-electron density matrix

$$\begin{aligned} \mathbf{D}^{\dagger} &= \mathbf{D} \\ \mathrm{Tr}\,\mathbf{DS} &= N_{\mathrm{elec.}} \\ \mathbf{DSD} &= \mathbf{D}\,, \end{aligned} \tag{B-6}$$

one straightforwardly obtains the following relations for $\mathbf{\Delta}$

$$\begin{aligned} \mathbf{\Delta}^{\dagger} &= \mathbf{\Delta} \\ \mathrm{Tr}\,\mathbf{\Delta S}^{-1} &= N_{\mathrm{elec.}} \\ \mathbf{\Delta S}^{-1}\mathbf{\Delta} &= \mathbf{\Delta}\,. \end{aligned} \tag{B-7}$$

Although Eqs. (B-6) and Eqs. (B-7) are formally equivalent, the equations for the standard AO density matrix $\mathbf{D}$ are somewhat simpler to use as they contain the metric $\mathbf{S}$ whereas the equations for $\mathbf{\Delta}$ involves the inverted metric $\mathbf{S}^{-1}$. It should be noted that Eqs. (B-7) are necessary and sufficient conditions, so all three equations are fulfilled if and only if $|0\rangle$ is a normalized single-determinant wave function.

Acknowledgements

A number of people have made my four years of PhD study a pleasant and interesting experience, and I could not have done it without them. First of all I would like to thank Jeppe Olsen and Poul Jørgensen for guidance and support through the years; they are a fantastic team. I am grateful to the whole theoretical chemistry group for nice lunch breaks and cake-meetings, and I would like to thank in particular Ove Christiansen for his career advices and Andreas Hesselman for sharing some of his latest work with me. And Stinne, how I managed to get through the days before Stinne joined the group is a mystery. It quickly turned out that we have much the same attitude towards life and we have shared many a wholehearted opinion of the life as such and our work situation in particular.

I would like to thank Pawel Salek for being good company during development and debugging of Fortran90 code of the finest quality and for being willing to help with any problems that I might have. A special thanks goes to Sonia Coriani and her husband Asger Halkier who took very good care of me during my visits in Trieste (even though I still havn't tasted her mum's lasagna).

For a number of conferences, winter schools and summer schools a group of mainly Scandinavian people made my trips an extra pleasant experience. They were always ready for some boozing and all sorts of crazy ideas. In particular should be mentioned Patzke-guy; a gentleman disguised as a theoretician, Pekka; the lizard king, Ulf; the sweet Swede, crazy Mikael, Ola, Tommy and all the others. It has been some really fine hours spent with you guys, and I hope to see you all again, maybe for a salmari or two – no miksi ei.

I also had the pleasure to spend a summer school with some of the students from the Copenhagen group: Marianne, Anders, Jacob and Thorsten. Anders and Jacob got connected to the Aarhus group at some point and have always been up for a nice chat and disgusting body noises to cheer up a grey day at work.

I would like to thank Birgit Schiøtt for nice colleagueship in connection with teaching and for coffee and talks in her office. I look forward to our collaboration on my next project.

I am grateful to the girl-gang; Louise, Trine, Cindie, and Rikke for keeping the connection to Århus and for gossip, lunch dates and girl nights.

I would also like to thank my parents for raising me as a good girl who always did her homework, otherwise I would never have gotten this far, and last but not least a great thanks goes to Kristoffer for putting up with me and being considerate and caring when needed.

References

[1] C. C. J. Roothaan, Rev. modern Physics **23**, 69 (1951).

[2] G. G. Hall, Proc. R. Soc. London, Ser. A **205**, 541 (1951).

[3] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[4] J. Koutecky and V. Bonacic, J. Chem. Phys. **55**, 2408 (1971); T. Claxton and W. Smith, Theor. Chim. Acta **22**, 399 (1971); W. A. Lathan, L. A. Curtiss, W. J. Hehre et al., *Progress in Physical Organic Chemistry*. (Wiley, New York, 1974).

[5] D. H. Sleeman, Theor. Chim. Acta **11**, 135 (1968).

[6] J. C. Slater, J. B. Mann, T. M. Wilson et al., Phys. Rev. **184**, 672 (1969); A. D. Rabuck and G. E. Scuseria, J. Chem. Phys. **110**, 695 (1999); B. I. Dunlap, Phys. Rev. A **29**, 2902 (1984).

[7] R. McWeeny, Proc. R. Soc. London Ser. A **235**, 496 (1956).

[8] R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).

[9] R. Fletcher and C. M. Reeves, Comput. J. **7**, 149 (1964).

[10] I. H. Hillier and V. R. Saunders, Proc. R. Soc. London Ser. A **320**, 161 (1970).

[11] R. Seeger and J. A. Pople, J. Chem. Phys. **65**, 265 (1976).

[12] R. N. Camp and H. F. King, J. Chem. Phys. **75**, 268 (1981).

[13] R. E. Stanton, J. Chem. Phys. **75**, 3426 (1981).

[14] W. R. Wessel, J. Chem. Phys. **47**, 3253 (1967); Douady, Ellinger, Subra et al., J. Chem. Phys. **72**, 1452 (1980).

[15] G. B. Bacskay, Chem. Phys. **61**, 385 (1981).

[16] R. Shepard, I. Shavitt, and J. Simons, J. Chem. Phys. **76**, 543 (1982).

[17] H. J. Aa. Jensen and P. Jørgensen, J. Chem. Phys. **80**, 1204 (1984); H. J. Aa. Jensen and H. Ågren, Chem. Phys. Lett. **110**, 140 (1984).

[18] X. Li, J. M. Millam, G. E. Scuseria et al., J. Chem. Phys. **119**, 7651 (2003); E. Hernández, M. J. Gillan, and C. M. Goringe, Phys. Rev. B **53**, 7147 (1996); J. M. Millam and G. E. Scuseria, J. Chem. Phys. **106**, 5569 (1997); M. Challacombe, J. Chem. Phys. **110**, 2332 (1999).

[19] A. H. R. Palser and D. E. Manolopoulos, Phys. Rev. B **58**, 12704 (1998).

[20] C. Ochsenfeld and M. Head-Gordon, Chem. Phys. Lett. **270**, 399 (1997).

[21] R. W. Nunes and D. Vanderbilt, Phys. Rev. B **50**, 17611 (1994); M. S. Daw, Phys. Rev. B **47**, 10895 (1993); X. P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).

[22] G. Galli and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992); F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993); W. Kohn, Chem. Phys. Lett. **208**, 167 (1993); P. Ordejon, D. Drabold, M. Grunbach et al., Phys. Rev. B **48**, 14646 (1993).

[23] T. Helgaker, H. Larsen, J. Olsen et al., Chem. Phys. Lett. **327**, 397 (2000).

[24] A. D. Daniels and G. E. Scuseria, Phys. Chem. Chem. Phys. **2**, 2173 (2000).

[25] J. VandeVondele and J. Hutter, J. Chem. Phys. **118**, 4365 (2003).

[26] J. B. Francisco, J. M. Martínez, and L. Martínez, J. Chem. Phys. **121**, 10863 (2004).

[27] D. R. Hartree, *The calculation of atomic structures*. (John Wiley and Sons, Inc., New York, 1957).

[28] E. Isaacson and H. B. Keller, *Analysis of numerical methods*. (Wiley, New York, 1966); C. C. J. Roothaan and P. S. Bagus, *Methods in Computational Physics*. (Academic, New York, 1963).

[29] N. W. Winter and T. H. Dunning Jr., Chem. Phys. Lett. **8**, 169 (1971).

[30] W. B. Neilsen, Chem. Phys. Lett. **18**, 225 (1973).

[31] M. C. Zerner and M. Hehenberger, Chem. Phys. Lett. **62**, 550 (1979).

[32] G. Karlström, Chem. Phys. Lett. **67**, 348 (1979).

[33] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980); P. Pulay, J. Comput. Chem. **3**, 556 (1982).

[34] H. Sellers, Int. J. Quant. Chem. **45**, 31 (1993).

[35] I. Hyla-Krispin, J. Demuynck, A. Strich et al., J. Chem. Phys. **75**, 3954 (1981).

[36] E. Cancès and C. Le Bris, Int. J. Quant. Chem. **79**, 82 (2000).

[37] K. N. Kudin, G. E. Scuseria, and E. Cancès, J. Chem. Phys. **116**, 8255 (2002).

[38] L. Thøgersen, J. Olsen, D. Yeager et al., J. Chem. Phys. **121**, 16 (2004).

[39] L. Thøgersen, J. Olsen, A. Köhn et al., J. Chem. Phys. **123**, 074103 (2005).

[40] A. P. Rendell, Chem. Phys. Lett. **229**, 204 (1994).

[41] H. Sellers, Chem. Phys. Lett. **180**, 461 (1991); C. Kollmar, Int. J. Quant. Chem. **62**, 617 (1997).

[42] V. R. Saunders and I. H. Hillier, Int. J. Quant. Chem. **7**, 699 (1973).

[43] S. P. Bhattacharyya, Chem. Phys. Lett. **56**, 395 (1978).

[44] R. Carbó, J. A. Hernández, and F. Sanz, Chem. Phys. Lett. **47**, 581 (1977).

[45] E. Cancès and C. Le Bris, Math. Model. Num. Anal. **34**, 749 (2000).

[46] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic Structure Theory*. (Wiley, Chichester, 2000).

[47] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).

[48] A. M. N. Niklasson, Phys. Rev. B **66**, 155115 (2002).

[49] E. Rubensson, Masters Thesis, Royal Institute of Technology (KTH), Stockholm, 2005.

[50] G. W. Stewart, *Introduction to Matrix Computations*. (Academic Press, inc., New York, 1973).

[51] J. W. Demmel, *Applied Numerical Linear Algebra*. (SIAM, 1997).

[52] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. (Wiley, New York, 1987).

[53] G. Chaban, M. W. Schmidt, and M. S. Gordon, Theor. Chem. Acc. **97**, 88 (1997); T. H. Fischer and J. E. Almlöf, J. Phys. Chem. **96**, 9768 (1992).

[54] R. E. Stanton, J. Chem. Phys. **75**, 5416 (1981).

[55] M. A. Natiello and G. E. Scuseria, Int. J. Quant. Chem. **26**, 1039 (1984).

[56] P. Cizek and J. Paldus, J. Chem. Phys. **47**, 3976 (1967); H. Fukutome, Int. J. Quant. Chem. **20**, 955 (1981); P. J. Thouless, Nucl. Phys. **21**, 225 (1960).

[57] V. Bach, E. H. Lieb, M. Loss et al., Phys. Rev. Lett. **72**, 2981 (1994); P.-L. Lions, Comm. Math. PHys. **109**, 33 (1987).

[58] L. E. Dardenne, N. Makiuchi, L. A. C. Malbouisson et al., Int. J. Quant. Chem. **76**, 600 (2000).

[59] A. Schafer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97**, 2571 (1992).

[60] A. Kalemos, T. H. Dunning Jr., and A. Mavridis, J. Chem. Phys. **123**, 014302 (2005); R. G. A. R. Maclagan and G. E. Scuseria, J. Chem. Phys. **106**, 1491 (1997); I. Shim and K. A. Gingerich, Int. J. Quant. Chem. **S23**, 409 (1989).

[61] H. Larsen, P. Jørgensen, J. Olsen et al., J. Chem. Phys. **113**, 8908 (2000).

[62] J. Olsen and P. Jørgensen, in *Modern Electronic Structure Theory, Part II*, edited by D. R. Yarkony (World Scientific, Singapore, 1995).

[63] J. Olsen and P. Jørgensen, J. Chem. Phys. **82**, 3235 (1985).

[64] J. Olsen, H. J. Aa. Jensen, and P. Jørgensen, J. Comp. Phys. **74**, 265 (1988).

[65] T. Helgaker and P. Jørgensen, Theor. Chim. Acta **75**, 111 (1989); T. Helgaker and P. Jørgensen, in *Advances in Quantum Chemistry* (Academic Press, 1988), Vol. 19; T. Helgaker and P. Jørgensen, in *Methods in Computational Molecular Physics*, edited by S. Wilson and G. H. F. Diercksen (Plenum Press, New York, 1992).

[66] H. Larsen, T. Helgaker, P. Jørgensen et al., J. Chem. Phys. **115**, 10344 (2001).

[67] D. Feller and J. A. Sordo, J. Chem. Phys. **113**, 485 (2000).

[68] D. Sherrill E. F. C. Byrd, and M. Head-Gordon, J. Phys. Chem. A **105**, 9736 (2001).

[69] J. Olsen, LUCIA, a quantum chemical program package.

[70] T. Helgaker, H. J. Aa. Jensen, P. Joergensen et al., DALTON, an electronic structure program (1997).

[71] T. H. Dunning Jr., J. Chem. Phys. **90**, 1007 (1989).

[72] K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecules*. (Van Nostrand, New York, 1979).

[73] W. Kutzelnigg, Theor. Chim. Acta **80**, 349 (1991).

[74] J. W. Krogh and J. Olsen, Chem. Phys. Lett. **344**, 578 (2001).

[75] L. Thøgersen and J. Olsen, Chem. Phys. Lett. **393**, 36 (2004).

[76] P. G. Szalay, L. Thøgersen, J. Olsen et al., J. Phys. Chem. A **108**, 3030 (2004).