NTNU

Norwegian University of
Science and Technology

Density-Based Formulation of Multi-Level Hartree-Fock Theory

Karen Oda Hjorth Dundas

Chemistry
Submission date: May 2016
Supervisor: Ida-Marie Høyvik, IKJ
Co-supervisor: Henrik Koch, IKJ

Norwegian University of Science and Technology
Department of Chemistry

Abstract

A scheme for multi-level (embedded) Hartree-Fock theory is developed. The goal is to reduce computational costs by treating different parts of an electronic system with different degrees of accuracy. In this thesis, a two-level scheme is considered, where one part is optimized through SCF iterations, and the other kept constant throughout optimization. Cholesky decomposition is used to partition the start guess density matrix. Test runs on some simple systems are presented, and show reasonably good agreement with otherwise equivalent non-embedded calculations, although we cannot at the present make any conclusions as to whether the method does in fact lower computational costs.

Sammendrag

Denne oppgaven er startfasen i utviklingen av *multi-level Hartree-Fock*(MLHF) teori, en modifikasjon av standard *Hartree-Fock teori* (HF) gjort for å redusere kostnadene ved beregninger på store elektroniske system. Tanken er å kunne dele opp et system i flere subsystem som så behandles med ulik grad av nøyaktighet. I denne oppgaven har vi holdt oss til to nivå, der ett optimeres via *self consistent field* (SCF) iterasjoner, mens den andre behandles som en konstant etter et initielt startgjett.

Vi forholder oss her til tetthetsbasert HF, og deler dermed den totale tetthetsmatrisen i to, en aktiv og en rest-del. Kun den aktive tettheten blir parametrisert og dermed oppdatert i løpet av iterasjonsmetoden, men man må ikke glemme at den totale energien (som er den vi er interessert i å studere), den totale gradienten og den totale Hessianen alle er avhengige av begge subsystemene. Dette gjør det nødvendig å utlede nye ligninger og å tilpasse implementeringen i LSDalton[1] til å passe vårt system.

Oppdelingen av startgjettet for den totale tetthetsmatrisen utføres ved Cholesky dekomponering, en metode som er vel egnet siden den konstruerer idempotente submatriser samtidig som den også bevarer denne egenskapen for den totale tetthetsmatrisa. Energien optimseres ved bruk av *andre-ordens Newton metode*, som løses i et redusert rom.

Test-beregninger er blitt utført med ulike enkle systemer. Som ventet ligger energiverdiene for MLHF høyere enn dem utført med fullt system, noe som følger direkte av variasjonsprinsippet, da vårt system er noe mer unøyaktig enn det totale. Samtidig er ikke avviket stort, og for tilstrekkelig store system kan man se for seg at denne feilen vil virke liten i forhold til hva man kan tjene på redusert data-kraft når metoden er ferdig utviklet.

Det er fremdeles en del problem som bør løses før denne metoden blir nyttig i praksis. I konstruksjonen av en idempotent matrise fra *Superposition of atomic densities* (SAD) startgjettet kreves diagonalisering av den totale Fock-matrisa, noe som er kostbart, spesielt for store system. Det bør her vurderes å innføres purifikasjon av tetthetsmatrisa istedet. Det bør også vurderes å benytte seg av approksimative metoder for Hessian informasjonen i Newton-likningene, som for eksempel *Augmented Roothan-Hall* (ARH) som unngår full konstruksjon av de lineære transformasjonene. Tilslutt må metoder for å konstruere et fornuftig level-shift implementeres, og prekondisjonering av Newton-likningene

må implementers.

Acknowledgements

First of all, I would like to thank my fantastic supervisor Ida-Marie Høyvik who has been of tremendous help and support throughout the last two years, and my co-supervisor Henrik Koch, for letting me work on this interesting project.

I also want to thank all the great people at my study hall with whom I have shared both good and bad times, and especially Linda Stakvik who read through my thesis and will forever be my number one LaTeX-guru.

Least but not last my love and gratitude goes to my parents Siv and Bjørn for always answering the phone to tell me I'm doing good or that I must remember to take care of myself. Without you, I would have been leading a miserable life the last two months.

Contents

Chapter 1

Introduction

A recurring issue within quantum chemistry is the question of accuracy versus computational cost. A variety of different methods have been developed, but the fact still remains: The more accurate our results, the more expensive our calculations. One of the earlier methods available is the *Hartree-Fock* (HF) method. For qualitative energy-calculation purposes, this is a good method, typically producing errors of around 1%[11]. Although many more sophisticated methods have been developed, such as *configuration interaction* (CI) or *coupled cluster* (CC)[5], wave functions generated through HF calculations are often used as initial start-points for these. HF is therefore still highly relevant, both in its own right, and as a preliminary step for more advanced calculations.

For large systems, calculations quickly get costly, and numerous attempts have been made to alleviate this. This thesis concerns one such attempt, the development of a new scheme to reduce the cost of Hartree-Fock calculations by treating different parts of a system with different degrees of accuracy, namely *multi-level*, or *embedded*, Hartree-Fock (MLHF). This is not a new idea, a variety of applications of multi-level schemes on different methods have been developed. Among them are procedures for *multi-level coupled cluster* (MLCC) calculations developed by Myhre et al.[23;24], embedding for *density functional theory* (DFT)[13], and the *IMOMO* method for embedded calculations of molecular orbitals by Humbel et al.[17]. Countless other examples can also be found. One of the main advantages of these sorts of schemes are that we can choose how we want to look at the respective parts of our total system. In certain situations, good empirical data might be available[13], and it might therefore be deemed unnecessary to recalculate these. In other instances, consider parts of the system might be considered best calculated with one method, while other methods better suit other parts.

Within Hartree-Fock there has not previously been conducted any significant work to develop a multi-level scheme, although a method called *Hartree-Fock embedded cluster* has been used by among others Miguel et al.[26] on inorganic calculations. The object of this thesis is to develop a density-based multi-level scheme for Hartree-Fock energy calculations. The idea is to first make a crude guess of the initial density matrix, and

then carry out optimization exclusively on the interesting part of the system.

Although it was stated in the previous paragraph that we start with a crude start guess, it is still of great importance that this initial density is not too inaccurate. It is true for any type of calculation that a bad start guess can ruin the whole calculation and impede convergence[29]. In MLHF, where optimization is only carried out on a small part of the system, we get the additional aspect that this bad start guess will be retained throughout calculations for the parts that are not optimized. The quality of the start guess is therefore of even greater importance in MLHF than in other schemes. There are many different ways to compute a start guess from a molecular geometry, among them the *Super-position of atomic densities* (SAD)[29] and the *extended Hückel method*[14;25]. We have used the former, and applied *Cholesky decomposition*[24;23] in order to correctly partition the system. In this way work can exclusively be done in the atomic orbital (AO) basis without having to actually compute the orbitals, only the density matrices.

When performing calculations in quantum chemistry there are a number of things we should be aware of in order to make our calculations as efficient as possible. Earlier, evaluation of Coulomb and exchange integrals were most time consuming[12], but ways to get around this have been found. Now, diagonalization of the Fock matrix has become on of the big obstacles. Helgaker et. al.[12] proposed *exponential parametrization* as a means to completely avoid diagonalization. However Newton equations in a reduces space still have to be solved, where the construction of the linear transformations of the Hessian on a trial vector ($H\mathbf{X}$) is a big obstacle. *Roothan-Hall* and *Augmented Roothan-Hall* methods[15] construct only an approximate version, but especially the latter still achieves good results.

Outline of thesis

In Chapter 2 elementary theory of quantum chemistry in general and Hartree-Fock theory in particular is reviewed. After looking at some founding principles, how the wave function is constructed and a quick introduction into Hartree-Fock theory, the density-based approach is introduced. Chapter 3 covers the development of the multi-level approach and how this affects the general equations from Chapter 2. In addition to this, certain things should be kept in mind when implementing these developments into LSDalton[1], and these aspects are discussed in Chapter 4. Results from test runs are presented in Chapter 5 and are discussed along with some other points in Chapter 6.

Chapter 2

Theory

Although it may appear, in many elementary textbooks on physical chemistry, as though quantum mechanical systems can be calculated exactly, this is not generally the case. When we look at even quite simple systems, we are forced to use approximative methods in order to obtain results. One of these methods is the Hartree-Fock method (HF), which is an improvement on the earlier *Self-Consistent Field* (SCF) method[5]. As all such methods it has advantages but also its drawbacks, and although it may no longer be counted among the main methods in quantum chemistry, its is still in use and is interesting in the sense that it is the foundation for many other more accurate methods[28]. In this chapter HF will be described in detail, but first certain foundations on which it is based will be introduced.

2.1 The Schrödinger equation

Perhaps the most important equation in quantum chemistry is the eigenvalue-equation used to calculate the energy E of electronic systems, the *Schrödinger equation*[4]

$$\hat{H}\psi = E\psi \tag{2.1.1}$$

where the *Hamiltonian operator* $\hat{H}$ contains the necessary terms for calculating the energy, such as the kinetic and potential contributions, and the *wave function* ψ is an electronic function of position and spin (see Section 2.2), and in certain situations, also time. The form of the Hamiltonian and the wave function are dependent on the specific system in question. The work of this thesis will be within the limits of the *Born Oppenheimer approximation* (BO)[4] where the nuclei are regarded as stationary in comparison with the electrons, as the former are so much heavier than the latter. Also this work only concerns itself with stationary systems, and therefore the wave function and Hamiltonian are only required to be dependent on position and spin. This enables us to write the time-independent electronic Hamiltonian (in atomic units)

$$\hat{H} = \sum_i \hat{h}(i) + \sum_{i>j} \hat{g}(i,j) + h_{nuc} \tag{2.1.2}$$

where $\hat{h}(i)$ and $\hat{g}(i,j)$ are defined as

$$\hat{h}(i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} \tag{2.1.3}$$

$$\hat{g}(i,j) = \frac{1}{r_{ij}} \tag{2.1.4}$$

Small indices sum over electrons, and capital indices sum over nuclei. Z_A is the charge of nucleus A, r_{iA} is the separation between electron i and nucleus A. r_{ij} is similarly the separation between electrons i and j.

Looking closer at $\hat{h}(i)$, we see that the first term is the electronic kinetic energy, while the second is the potential energy as a result of attractive forces between electrons and the nuclei. $\hat{g}(i,j)$ is the repulsive potential energy between the electrons. The last term, h_{nuc} is the potential energy caused by interactions between the nuclei, and it follows from BO that it can be treated as a constant in the system considered in this thesis.

Before proceeding, it should be noted that there are two rivalling notations used in quantum chemistry, *first-* and *second quantization*. First-quantization is the intuitive notation where we deal with standard eigenvalue problems, with operators working on functions. In second quantization, on the other hand, the eigenfunctions themselves are made up of operators, which will in turn alter the way the original operators are defined. They both have their advantages, but in this thesis it is sufficient to work in first-quantization only.

2.2 The electronic wave function

As mentioned in Section 2.1, to correctly describe a system, the wave function should be chosen to depend on both spatial and spin-coordinates (r and m_s respectively). This is the case for *spin-orbitals* ϕ_i, but not for *spatial* ψ_i, which are only dependent on position. When constructing spin-orbitals, it is convenient to make them products of a spatial orbital and a spin-function σ_i

$$\phi_i(x) = \psi_i(r)\sigma_i(m_s) \tag{2.2.1}$$

For electrons, the only allowed value of the spin quantum number s is $\frac{1}{2}$, but depending on the orientation of this spin, it can have spin magnetic quantum numbers m_s of either $\frac{1}{2}$ or $-\frac{1}{2}$[4]. These two states are often referred to as α- and β-spin.

There are two principal ways to combine spatial orbitals and spin-functions, either

restricted or *unrestricted*. One commonly used method is to require that pairs of electrons have the same spatial orbital, but opposite spin. This is called restricted HF (RHF), and is convenient for closed-shell systems. In a system containing N_e electrons, there will therefore be $\frac{N_e}{2}$ spatial orbitals, two spin-functions and N_e spin orbitals. The question now arises as to what we should do with open-shell systems. We can either keep the restricted approach to all the closed inner shells (*Restricted open-shell*), or allow all spinorbitals to have different spatial components (*Unrestricted open-shell Hartree-Fock*, UHF). The first approach is convenient, but has the disadvantage that we might miss out on certain exchange interactions caused by spin, which will give us a too high ground state energy. The second approach, UHF, gives better results for the energy through variation, but does not produce a total ground state wave function that is an eigenfunction of the total spin angular momentum[5;11]. This is, however, normally not a large problem[5].

The *Pauli principle*[4] states that the wave function must be antisymmetric in the sense that when we exchange two labels, the wave function must change sign. An example of this would be if the electrons of two orbitals are exchanged. To ensure that this is the case, spin-orbitals are combined in an antisymmetric way[21], typically as *Slater determinants* Φ[5]. These are normalized, antisymmetrized combinations of restricted spinorbitals and have the following form

$$\Phi(1,2,\ldots,N) = \left(\frac{1}{N!}\right)^{\frac{1}{2}} \begin{vmatrix} \phi_a(1) & \phi_b(1) & \ldots & \phi_z(1) \\ \phi_a(2) & \phi_b(2) & \ldots & \phi_z(2) \\ \vdots & \vdots & \ldots & \vdots \\ \phi_a(N) & \phi_b(N) & \ldots & \phi_z(N) \end{vmatrix} \tag{2.2.2}$$

2.3 The variational theorem

Another central theorem in quantum chemistry is the *variational theorem*[2] which states that for any guess made of a normalized wave function ψ^{trial}, the computed energy $E[\psi^{trial}]$ of the system will always be higher than the true ground state energy E_0 of the system[5;27]. This is given by the *Rayleigh ratio*

$$E[\psi^{trial}] \equiv \frac{\langle \psi^{trial} | \hat{H} | \psi^{trial} \rangle}{\langle \psi^{trial} | \psi^{trial} \rangle} \geq E_0 \tag{2.3.1}$$

The significance of (2.3.1) is that the lower (normally the more negative) we manage to make $E[\psi^{trial}]$, the nearer it is to E_0. Equality occurs only when ψ^{trial} is the true ground state wave function. Therefore, in order to make $E[\psi^{trial}]$ a better approximation of the ground state energy, we look at what parameters it contains and vary these in order to minimize it. This is done until we are satisfied with our result, and we have now found an upper bound to E_0.

As seen in the Section 2.2, a typical way to represent the wave function is with a Slater Determinant, which is an anitsymmetric combination of molecular spinorbitals ϕ_i. In this thesis, we will work only in the framework of RHF, which means that we only need to work with doubly occupied spatial orbitals ψ_i. These spatial MOs are typically constructed by linear combination of spatial atomic orbitals (AOs) χ_μ

$$\psi_i = \sum_\mu C_{\mu i}\chi_\mu \tag{2.3.2}$$

When $\langle\Phi|\hat{H}|\Phi\rangle$ is now minimized, it is with the object to find the set of coefficients $C_{\mu i}$ that minimize the energy.

It should be noted that a small error in the eigenvalues does not automatically imply a correspondingly small error in the eigenvectors[6]. This has to do with the shape of the function, and that it for the most part is important that approximations are good close to the nucleus. In other words, an eigenfunction that is good only at short distances might be preferable to one that is overall better, but is not as good at short range. At the same time, it is generally true that approximation for the ground state energy approaches the true value of E_0 much faster than ψ approaches the true ground state wave function[21]. So as long as we have half-good approximations to our eigenfunctions, we can still get good results for the energy-eigenvalues. Another point is that although a wave functions works well to calculate the energy, it might not produce a good result for other molecular properties[5].

2.4 Hartree-Fock theory

We are now ready to start discussing the main method of this thesis, namely the Hartree-Fock method (HF), an ab initio method which is an improvement of the earlier self-consistent field (SCF) method developed by Hartree[21]. Common to these methods is the average treatment of the electronic potential[5], but HF constructs it in a more accurate manner. For each electron this potential is a spherically treated average stemming from the $n-1$ other electrons, which greatly simplifies calculations. As we shall see later, the operators used in this method are dependent on its eigenfunctions, making HF iterative by necessity. Calculations are run until self-consistency is reached.

As explained in Section 2.2, spatial orbitals and spin functions are combined in such a way that all occupied spatial orbitals are doubly occupied, with electrons of opposite spin. We will for the remainder of this thesis be working in this closed-shell molecular orbital (MO) basis. From this point on, we will also restrict ourselves to only work with real eigenfunctions, $\psi^* = \psi$.

Looking back at (2.1.2), the Hamiltonian is made up of two operators, $\hat{h}$ and $\hat{g}$. Often however, it is more convenient to work with matrices, and we now define

$$\begin{aligned} h_{pq} &= \int \psi_p(r_1)\hat{h}(1)\psi_q(r_1)dr_1 \\ &= \langle p|\hat{h}|q\rangle \end{aligned} \tag{2.4.1}$$

$$g_{pqrs} = \int \phi_p(r_1)\phi_q(r_1)\hat{g}(1,2)\phi_r(r_2)\phi_s(r_2)dr_1dr_2 \tag{2.4.2}$$

for electron 1 and 2. Up until now, electrons have been labeled with the indices i, j, but from this point on, indices will only be needed for orbitals. The indices $p, q, r, s, ...$ are therefore defined to label MOs in general, $i, j, k, l, ...$ to label occupied MOs and $a, b, c, d, ...$ to label virtual MOs. The Greek labels μ, ν, ρ, σ still label AOs. Applying this new notation, the definitions (2.4.1) and (2.4.2) and the fact that all occupied orbitals are doubly occupied spatial orbitals, the Hartree-Fock energy can be written as

$$E_{HF} = 2\sum_{i}^{N_e/2} h_{ii} + \sum_{ij}^{N_e/2}(2g_{iijj} - g_{ijji}) + h_{nuc} \tag{2.4.3}$$

In (2.4.3), the potential is made up of to parts, g_{iijj} and g_{ijji}. These are called, respectively, the *Coulomb integral* and the *Exchange integral*.

As presented in Section 2.3, the MOs are constructed as a linear combination of AOs, and the coefficients are optimized to find the minimum energy. To do this, a Lagrangian with the restriction that spatial MOs are orthonormal are introduced(Helgaker et al.[11])

$$\begin{aligned} L(\mathbf{C}) =& E(\mathbf{C}) - 2\sum_{ij}\varepsilon_{ij}\left(\langle\psi_i|\psi_j\rangle - \delta_{ij}\right) \\ =& E(\mathbf{C}) - 2\sum_{ij}\varepsilon_{ij}\left(\sum_{\mu\nu} C_{\mu i}S_{\mu\nu}C_{\nu j} - \delta_{ij}\right) \end{aligned} \tag{2.4.4}$$

ε_{ij} is a Lagrangian multiplier, but we shall later see that it is also the orbital energies. $S_{\mu\nu}$ are the matrix elements of the *overlap matrix*, a measure of the overlap between the AOs

$$S_{\mu\nu} = \langle\chi_\mu|\chi_\nu\rangle \tag{2.4.5}$$

The next step is to differentiate (2.4.4) with respect to the coefficients and setting the resulting equation equal to zero. This leads us to the Hartree-Fock equations

$$\mathbf{fC} = \mathbf{SC}\boldsymbol{\varepsilon} \tag{2.4.6}$$

$\boldsymbol{\varepsilon}$ is a diagonal matrix with orbital energies on the diagonal. The *Fock matrix* $\mathbf{f}$ is the

mean-field Hamiltonian containing 1-electron interactions and the averagely treated 2-electron interactions with elements

$$f_{\mu\nu} = h_{\mu\nu} + \sum_i (2g_{\mu\nu ii} - g_{\mu ii\nu}) \tag{2.4.7}$$

From (2.4.7) we see that the Fock matrix is dependent on the exact same coefficients that are solved for in (2.4.6). SCF is therefore required. Solving the Hartree-Fock equations (2.4.6) yields a diagonalization of the Fock matrix, we however do not need to restrict ourselves to a diagonalized Fock matrix. A block diagonal one is sufficient since the optimization condition only requires the occupied-virtual Fock matrix blocks to be zero. As we shall see in Section 2.5, we can from now on work entirely in the AO basis.

2.5 Density-based Hartree-Fock theory

In the previous section we looked at the Hartree-Fock equations as functions of the MO coefficients used to combine atomic spatial orbitals into molecular spatial orbitals

$$\psi_i = \sum_\mu \chi_\mu C_{\mu i} \tag{2.5.1}$$

or equivalently

$$|i\rangle = \sum_\mu |\mu\rangle C_{\mu i} \tag{2.5.2}$$

From here on we will only be working with *density matrices*, that describe the electronic occupation of atomic orbitals. The elements of the atomic density matrix can be written as

$$\begin{aligned} D_{\mu\nu} =& 2\sum_i C_{\mu i} C_{\nu i} \\ =& (\mathbf{C}\mathbf{D}^{MO}\mathbf{C}^T)_{\mu\nu} \end{aligned} \tag{2.5.3}$$

where the molecular density matrix $\mathbf{D}^{MO}$ is diagonal with diagonal values of 2 for the $N_e/2$ occupied spatial molecular orbitals, and zero for all other elements

$$\mathbf{D}^{MO} = 2\begin{bmatrix} \mathbf{1}_{N_e/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{2.5.4}$$

The next goal is to write the energy in terms of $\mathbf{D}$. Using the bracket notation introduced in (2.4.1) we can now write ($\mathbf{h}$ is symmetric)

$$\begin{aligned}\sum_i h_{ii} =&\langle i|\hat{h}|i\rangle = \sum_i \sum_{\mu\nu} \langle \mu|\hat{h}|\nu\rangle C_{\mu i} C_{\nu i} = \frac{1}{2}\sum_{\mu\nu} h_{\mu\nu} D_{\mu\nu} \\ =&\frac{1}{2}\sum_{\nu} (\mathbf{hD})_{\nu\nu} = \frac{1}{2} Tr(\mathbf{hD})\end{aligned} \tag{2.5.5}$$

and

$$\begin{aligned}&\sum_{ij} (2g_{iijj} - g_{ijji}) \\ =&\sum_{ij}\left[2\sum_{\mu\nu\rho\sigma} C_{\mu i} C_{\nu i} C_{\rho j} C_{\sigma j} g_{\mu\nu\rho\sigma} - \sum_{\mu\nu\rho\sigma} C_{\mu i} C_{\nu j} C_{\rho j} C_{\sigma i} g_{\mu\nu\rho\sigma}\right] \\ =&\frac{1}{2}\sum_{\mu\nu\rho\sigma} D_{\mu\nu} D_{\rho\sigma} g_{\mu\nu\rho\sigma} - \frac{1}{4}\sum_{\mu\nu\rho\sigma} D_{\mu\sigma} D_{\nu\rho} g_{\mu\nu\rho\sigma} \\ =&\frac{1}{4}\sum_{\mu\nu\rho\sigma} (2g_{\mu\nu\rho\sigma} - g_{\mu\sigma\rho\nu}) D_{\mu\nu} D_{\rho\sigma}\end{aligned} \tag{2.5.6}$$

From now on, all quantities where the basis is not otherwise specified should be assumed to be in the AO basis. To simplify notation, the matrix $\mathbf{G}$ for the two-electron interactions is introduced

$$G_{\mu\nu}(A) = \sum_{\rho\sigma} (2g_{\mu\nu\rho\sigma} - g_{\mu\sigma\rho\nu}) A_{\rho\sigma} \tag{2.5.7}$$

which gives the total equation

$$\begin{aligned}E =&2\sum_i h_{ii} + \sum_{ij} (2g_{iijj} - g_{ijji}) + h_{nuc} \\ =&Tr(\mathbf{hD}) + \frac{1}{4} Tr(\mathbf{DG(D)}) + h_{nuc}\end{aligned} \tag{2.5.8}$$

In Sections 2.4 we talked about the Fock operator which is the (one-electron) operator in Hartree-Fock equations. In the density-based formulation, the Fock matrix is defined as

$$f_{\mu\nu} = h_{\mu\nu} + \frac{1}{2}\sum_{\rho\sigma} D_{\rho\sigma} \left(2g_{\mu\nu\rho\sigma} - g_{\mu\sigma\rho\nu}\right) \tag{2.5.9}$$

which gives us

$$2\left(f_{\mu\nu} - h_{\mu\nu}\right) = \sum_{\rho\sigma} D_{\rho\sigma}\left(2g_{\mu\nu\rho\sigma} - g_{\mu\sigma\rho\nu}\right) \tag{2.5.10}$$

Inserting this into (2.5.8) the energy becomes

$$\begin{aligned} E =& \sum_{\mu\nu} h_{\mu\nu} D_{\mu\nu} + \frac{1}{2}\sum_{\mu\nu} D_{\mu\nu}\left(f_{\mu\nu} - h_{\mu\nu}\right) + h_{nuc} \\ =& \frac{1}{2} Tr(\mathbf{Dh}) + \frac{1}{2} Tr(\mathbf{Df}) + h_{nuc} \end{aligned} \tag{2.5.11}$$

2.6 Parametrization of transformations on the AO density

For a given electronic state there are many possible determinants and therefore also many possible density matrices(Helgaker et al.[11]) and MO coefficients.

In order to find these other possible coefficients and density matrices, unitary transformations are performed on the coefficients through the parametrization $\tilde{C} = C\exp(\boldsymbol{\kappa})$, where $\boldsymbol{\kappa}$ is an antisymmetric matrix that varies our system, in the MO basis. As previously stated, this thesis concerns work done in the AO basis, we therefore define a corresponding antisymmetric matrix in the AO basis, $\mathbf{X}$, related to $\boldsymbol{\kappa}$ as $\mathbf{X} = \mathbf{C}\boldsymbol{\kappa}\mathbf{C}^T$.

Before proceeding, it should be noted that the density matrix considered until now requires diagonalization. If possible, this should be avoided, and a new density matrix in the AO basis, $\mathbf{R}$, is therefore introduced. It is a scaled version og the normal density matrix $\mathbf{D}$, so that $\mathbf{R} = \frac{1}{2}\mathbf{D}$. $\mathbf{R}$ does not require diagonalization, and it fullfills the conditions

$$\mathbf{R}^T = \mathbf{R} \tag{2.6.1}$$

$$Tr\mathbf{RS} = \frac{1}{2}N \tag{2.6.2}$$

$$\mathbf{RSR} = \mathbf{R} \tag{2.6.3}$$

In particular the last equation, (2.6.3), the *Idempotency condition* is of great importance in this work. All of the equations (2.6.1), (2.6.2) and (2.6.3) are general properties for scaled density matrices for closed-shell systems with doubly occupied orbitals. In order to vary $\mathbf{R}$ we parametrize it with $\mathbf{X}$

$$\mathbf{R}(\mathbf{X}) = \exp(-\mathbf{XS})\mathbf{R}\exp(\mathbf{SX}) \tag{2.6.4}$$

which can be expanded as

$$\mathbf{R}(\mathbf{X}) = \mathbf{R} + [\mathbf{R}, \mathbf{X}]_S + \frac{1}{2}[[\mathbf{R}, \mathbf{X}]_S, \mathbf{X}]_S + \dots \tag{2.6.5}$$

When using this in further equations, we will truncate after second order in $\mathbf{X}$.

2.7 Electron correlation

In Section 2.4 it was seen that one of the main advantages with HF is the average way it treats the electronic potential. This, however, is also the method's biggest problem. It disregards instantaneous electronic interactions and the "quantum mechanical effects of the electron distribution" [5], in other words it ignores *electron correlation*.

This stems from the fact that the potential is treated in an average way, as a spherical potential surrounding the electron in question [5]. Instantaneous interactions between the $n-1$ other electrons are completely disregarded. This is a problem as the electrons tend to stay away from each other because of the repulsive forces between them. In a true system, the probability of finding two electrons, and especially two with the same spin, close by each other is very low. The HF method avoids electrons with the same spin being close to each other, but doesn't set restrictions for those of opposite spin. The difference between the Hartree-Fock energy and the true nonrelativistic energy is called the *correlation energy*.

$$E_{corr} = E_{exact} - E_{HF} \tag{2.7.1}$$

The most obvious way to solve this is by including terms of inter-electronic distances, but this becomes complicated for atoms with more than a few electrons. As a result, a multitude of post-HF methods have been developed.

In the HF method described above, our wave function only includes the ground state. However, in *configuration state functions* (CSFs), a combination of excited states are included as well [5]. The *configuration interaction* (CI) method constructs its wave function as a linear combination of CSFs. In HF calculations, our accuracy is always dependent on the number of basis functions we include. This is true also for CI, but in addition we are limited by the number of CSFs we construct from this given basis set. The case where all "CSFs of appropriate symmetry are included for a given finite basis set" [5] is called *full CI*. This obviously becomes costly quite quickly, but results from full CI are often used as a reference point in comparison with other methods [11].

In CI calculations the coefficients $c_{\mu i}$ determining the combination of basis function in each spatial orbital are predetermined with HF and kept fixed, it is only the coefficients of the excited determinants that are optimized. An extended method where also $c_{\mu i}$ are varied is the *multiconfiguration self-consistent field method* (MCSCF) [5]. By definition this method will be quite expensive, but it allows us to use a smaller number of CSFs than

in CI[5].

The two previous methods are both variational, but not size extensive. Common size-extensive methods are *perturbation theory* and *coupled cluster theory* (CC).

Before concluding this discussion about alternative methods, *density functional theory* (DFT) ought to be mentioned. This, unlike the ones discussed above, can not be called a post HF method, as it builds on a completely different concept than the ab inito approaches. The idea is that all properties of the electronic system can be described by the position-dependent electronic distribution[5]. A variety of semi-empirical methods are also available.

Chapter 3

Multi-level Hartree-Fock theory

As mentioned in the Introduction, this thesis is a study of a system that is such that not all its parts have to be calculated to the same degree of accuracy. This is called multilevel, or embedded, Hartree-Fock, where the word multilevel comes from the fact that we will be looking at different parts of the system with different levels of accuracy. In this thesis, only two levels will be considered. The total system is therefore divided into two parts, one that is to be calculated accurately (the *active part*, $\mathbf{D}_a$), and one that can be viewed in a more crude manner (the *rest-part* or *inactive part*, $\mathbf{D}_r$). Each of these will have its own density matrix, and the sum of these two is the density matrix for the whole system

$$\mathbf{D} = \mathbf{D}_a + \mathbf{D}_r \tag{3.0.1}$$

The following equations are all based on the ones presented in Chapter 2, which are all from Helgaker et al.[12;11]. The first sections of this chapter covers the fitting of the equations to our multilevel system. We will also derive how the equations for the linear transformations will look in our new scheme and conclude with a quick note about preservation of idempotency.

3.1 Ground state energy for the multi-level scheme

Starting off with (2.5.8), we fit it to the multi-level scheme by inserting (3.0.1). Using that $Tr(\mathbf{D}_r\mathbf{G}(\mathbf{D}_a)) = Tr(\mathbf{D}_a\mathbf{G}(\mathbf{D}_r))$ this gives

$$\begin{aligned} E =& Tr((\mathbf{D}_a + \mathbf{D}_r)\mathbf{h}) + \frac{1}{4}Tr((\mathbf{D}_a + \mathbf{D}_r)\mathbf{G}(\mathbf{D})) + h_{nuc} \\ =& Tr(\mathbf{D}_a\mathbf{h}) + Tr(\mathbf{D}_r\mathbf{h}) + \frac{1}{4}Tr(\mathbf{D}_a\mathbf{G}(\mathbf{D}_a)) \\ &+ \frac{1}{4}Tr(\mathbf{D}_r\mathbf{G}(\mathbf{D}_r)) + \frac{1}{2}Tr(\mathbf{D}_r\mathbf{G}(\mathbf{D}_a)) + h_{nuc} \end{aligned} \tag{3.1.1}$$

As repeatedly stated throughout this thesis, the Fock matrix plays an important role in quantum chemistry. We now want to write an equivalent of (2.5.11) with the split-density notation. Since the Fock matrix is dependent upon $\mathbf{D}$, it is necessary to define one Fock matrix for the active part and one for the rest-part. For the active part, define

$$f_{a,\mu\nu} = h_{\mu\nu} + \frac{1}{2}\sum_{\rho\sigma} D^{AO}_{a,\rho\sigma}\left(2g_{\mu\nu\rho\sigma} - g_{\mu\sigma\rho\nu}\right) \tag{3.1.2}$$

and equivalently for the rest-part. The Fock matrix of the total system is then

$$f_{\mu\nu} = f_{a,\mu\nu} + f_{r,\mu\nu} - h_{\mu\nu} \tag{3.1.3}$$

Inserting this into (2.5.11), we get

$$\begin{aligned} E =& \frac{1}{2}\sum_{\mu\nu} D_{\mu\nu}h_{\mu\nu} - \frac{1}{2}\sum_{\mu\nu} D_{\mu\nu}h_{\mu\nu} + \frac{1}{2}\sum_{\mu\nu} f_{a,\mu\nu}D_{\mu\nu} + \frac{1}{2}\sum_{\mu\nu} f_{r,\mu\nu}D_{\mu\nu} + h_{nuc} \\ =& \frac{1}{2}\sum_{\mu\nu} f_{a,\mu\nu}D_{\mu\nu} + \frac{1}{2}\sum_{\mu\nu} f_{r,\mu\nu}D_{\mu\nu} + h_{nuc} \\ =& \frac{1}{2}Tr(\mathbf{D}_a\mathbf{f}_a) + \frac{1}{2}Tr(\mathbf{D}_a\mathbf{f}_r) + \frac{1}{2}Tr(\mathbf{D}_r\mathbf{f}_a) + \frac{1}{2}Tr(\mathbf{D}_r\mathbf{f}_r) + h_{nuc} \end{aligned} \tag{3.1.4}$$

3.2 Parametrization of the ground state energy

As mentioned in Section 2.6, the density matrix $\mathbf{D}$ is inconvenient to use in calculations. It is therefore more common to use its scaled version, $\mathbf{R} = \frac{1}{2}\mathbf{D}$, which is split into an active part and a rest-part just as for the original unscaled system, see (3.0.1). If we now define

$$\mathbf{R}_a = \frac{1}{2}\mathbf{D}_a \tag{3.2.1}$$

$$\mathbf{R}_r = \frac{1}{2}\mathbf{D}_r \tag{3.2.2}$$

we have that $\mathbf{R} = \mathbf{R}_a + \mathbf{R}_r$ which is analogous to (3.0.1).

The main objective of this thesis is the energy-optimization of the active part of our system. In order to do this we need to find the first- and second-order derivatives of this energy. To be able to vary the system, the active scaled density matrix $\mathbf{R}_a$ is parametrized as in Section 2.6, and the energy expression is differentiated with respect to $\mathbf{X}$. Since $\mathbf{R}_r$ does not depend on $\mathbf{X}$, all the the energy-terms dependent only on $\mathbf{R}_r$ will disappear upon differentiation, and can therefore be eliminated from our energy-equation in the situations where we are only interested in finding its derivatives. We write $(E - E_r)$ as (3.1.1), minus the pure rest terms and insert (3.2.2) and the expanded version of (3.2.1)

(as in (2.6.5)) to get

$$\begin{aligned}(E-E_r)(\mathbf{X}) =& E^{(0)} + 2\sum_{\mu\nu}([\mathbf{R}_a,\mathbf{X}]_S)_{\mu\nu}h_{\mu\nu} + \sum_{\mu\nu}([[\mathbf{R}_a,\mathbf{X}]_S,\mathbf{X}]_S)_{\mu\nu}h_{\mu\nu} \\ &+\sum_{\mu\nu}(\mathbf{R}_a)_{\mu\nu}(\mathbf{G}([\mathbf{R}_a,\mathbf{X}]_S))_{\mu\nu} + \frac{1}{2}\sum_{\mu\nu}(\mathbf{R}_a)_{\mu\nu}(\mathbf{G}([[\mathbf{R}_a,\mathbf{X}]_S,\mathbf{X}]_S))_{\mu\nu} \\ &+\sum_{\mu\nu}([\mathbf{R}_a,\mathbf{X}]_S)_{\mu\nu}(\mathbf{G}(\mathbf{R}_a))_{\mu\nu} + \sum_{\mu\nu}([\mathbf{R}_a,\mathbf{X}]_S)_{\mu\nu}(\mathbf{G}([\mathbf{R}_a,\mathbf{X}]_S))_{\mu\nu} \\ &+\frac{1}{2}\sum_{\mu\nu}([[\mathbf{R}_a,\mathbf{X}]_S,\mathbf{X}]_S)_{\mu\nu}(\mathbf{G}(\mathbf{R}_a))_{\mu\nu} \\ &+2\sum_{\mu\nu}([\mathbf{R}_a,\mathbf{X}]_S)_{\mu\nu}(\mathbf{G}(\mathbf{R}_r))_{\mu\nu} + \sum_{\mu\nu}([[\mathbf{R}_a,\mathbf{X}]_S,\mathbf{X}]_S)_{\mu\nu}(\mathbf{G}(\mathbf{R}_r))_{\mu\nu} \\ &+O(\mathbf{X}^3) \end{aligned} \tag{3.2.3}$$

$E^{(0)}$ contains all terms that are zero-order in $\mathbf{X}$ and $O(\mathbf{X}^3)$ contains all the higher-order terms. The reason that for not expanding $\mathbf{R}_r$ is that we view this as a constant matrix, and therefore do not want to optimize it by varying $\mathbf{X}$.

For two symmetrical matrices $\mathbf{A}$ and $\mathbf{B}$ we have the relation[11]

$$Tr(\mathbf{A}\mathbf{G}(\mathbf{B})) = Tr(\mathbf{B}\mathbf{G}(\mathbf{A})) \tag{3.2.4}$$

which for us has the nice consequence that

$$Tr(\mathbf{R}_a\mathbf{G}([\mathbf{R}_a,\mathbf{X}]_S)) = Tr([\mathbf{R}_a,\mathbf{X}]_S\,\mathbf{G}(\mathbf{R}_a)) \tag{3.2.5}$$

$$Tr(\mathbf{R}_a\mathbf{G}([[\mathbf{R}_a,\mathbf{X}]_S,\mathbf{X}]_S) = Tr([[\mathbf{R}_a,\mathbf{X}]_S,\mathbf{X}]_S\,\mathbf{G}(\mathbf{R}_a)) \tag{3.2.6}$$

It is desirable to exchange the terms with $\mathbf{G}(\mathbf{R}_a)$ with terms containing the Fock matrix. The active Fock matrix is

$$\mathbf{f}_a = \mathbf{h} + \mathbf{G}(\mathbf{R}_a) \tag{3.2.7}$$

Inserting this into (3.2.3) gives

$$\begin{aligned}(E-E_R)(\mathbf{X}) =& E^{(0)} + 2Tr(\mathbf{f}_a\,[\mathbf{R}_a,\mathbf{X}]_S) + Tr(\mathbf{f}_a\,[[\mathbf{R}_a,\mathbf{X}]_S,\mathbf{X}]_S) \\ &+Tr([\mathbf{R}_a,\mathbf{X}]_S\,\mathbf{G}([\mathbf{R}_a,\mathbf{X}]_S)) + 2Tr([\mathbf{R}_a,\mathbf{X}]_S\,\mathbf{G}(\mathbf{R}_r)) \\ &+Tr([[\mathbf{R}_a,\mathbf{X}]_S,\mathbf{X}]_S\,\mathbf{G}(\mathbf{R}_r)) + O(\mathbf{X}^3) \end{aligned} \tag{3.2.8}$$

A useful trait of the matrix $\mathbf{X}$ is that it can be written in terms of $\mathbf{E}^-_{\mu\nu}$[11], the anit-

symmetric combination of the elementary matrices $\mathbf{E}_{\mu\nu}$ (refer to Helgaker et al.[11] for definition)

$$\mathbf{E}^{-}_{\mu\nu} = \mathbf{E}_{\mu\nu} - \mathbf{E}_{\nu\mu} \tag{3.2.9}$$

in the following manner

$$\mathbf{X} = \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^{-}_{\mu\nu} \tag{3.2.10}$$

This enables us to write (3.2.8) as

$$\begin{aligned}
(E - E_R)(\mathbf{X}) =& E^{(0)} + 2Tr\left(\mathbf{f}_a\left[\mathbf{R}_a, \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^{-}_{\mu\nu}\right]_S\right) \\
&+ Tr\left(\mathbf{f}_a\left[\left[\mathbf{R}_a, \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^{-}_{\mu\nu}\right]_S, \sum_{\rho>\sigma} X_{\rho\sigma}\mathbf{E}^{-}_{\rho\sigma}\right]_S\right) \\
&+ Tr\left(\left[\mathbf{R}_a, \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^{-}_{\mu\nu}\right]_S \mathbf{G}\left(\left[\mathbf{R}_a, \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^{-}_{\mu\nu}\right]_S\right)\right) \\
&+ 2Tr\left(\left[\mathbf{R}_a, \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^{-}_{\mu\nu}\right]_S \mathbf{G}(\mathbf{R}_r)\right) \\
&+ Tr\left(\left[\left[\mathbf{R}_a, \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^{-}_{\mu\nu}\right]_S, \sum_{\rho>\sigma} X_{\rho\sigma}\mathbf{E}^{-}_{\rho\sigma}\right]_S \mathbf{G}(\mathbf{R}_r)\right) + O(\mathbf{X}^3)
\end{aligned} \tag{3.2.11}$$

This will now be used to find the electronic gradient and Hessian.

3.3 Hartree-Fock electronic gradient and Hessian

Both the electronic gradient and Hessian are evaluated at $\mathbf{X} = 0$, which means that to find the gradient, only the terms that are first order in $\mathbf{X}$ are needed, and for the Hessian only those that are second order in $\mathbf{X}$ needed. Also, since only the (scaled) density matrix for the active part is a function of $\mathbf{X}$, it is unproblematic to leave the terms only dependent on the rest-part (E_r) out of the calculations. The terms from (3.2.11) needed to find the HF gradient are therefore

$$
\begin{aligned}
E_{1,\mu\nu}(\mathbf{X}) =& 2Tr\left(\mathbf{f}_a\left[\mathbf{R}_a, X_{\mu\nu}\mathbf{E}^-_{\mu\nu}\right]_S\right) \\
&+2Tr\left(\left[\mathbf{R}_a, X_{\mu\nu}\mathbf{E}^-_{\mu\nu}\right]_S \mathbf{G}(\mathbf{R}_r)\right)
\end{aligned} \tag{3.3.1}
$$

(3.3.1) is now differentiated with respect to $\mathbf{X}$ and the cyclic trace rule is used in order to apply certain features of $\mathbf{E}^-_{\mu\nu}$ and $\mathbf{E}_{\mu\nu}$[11]. First, for all antisymmetric $\mathbf{A}$

$$
Tr(\mathbf{E}^-_{\mu\nu}\mathbf{A}) = 2Tr(\mathbf{E}_{\mu\nu}\mathbf{A}) \tag{3.3.2}
$$

and secondly, for all $\mathbf{M}$

$$
Tr(\mathbf{E}_{\mu\nu}\mathbf{M}) = M_{\nu\mu} \tag{3.3.3}
$$

Using all this, the electronic gradient becomes

$$
\begin{aligned}
E^{(1)}_{\mu\nu}(\mathbf{X}) =& 4\left(\mathbf{f}_a\mathbf{R}_a\mathbf{S} - \mathbf{S}\mathbf{R}_a\mathbf{f}_a\right)_{\nu\mu} \\
&+4\left(\mathbf{G}(\mathbf{R}_r)\mathbf{R}_a\mathbf{S} - \mathbf{S}\mathbf{R}_a\mathbf{G}(\mathbf{R}_r)\right)_{\nu\mu} \\
=& 4\left(\mathbf{S}\mathbf{R}_a\mathbf{f}_a - \mathbf{f}_a\mathbf{R}_a\mathbf{S}\right)_{\mu\nu} \\
&+4\left(\mathbf{S}\mathbf{R}_a\mathbf{G}(\mathbf{R}_r) - \mathbf{G}(\mathbf{R}_r)\mathbf{R}_a\mathbf{S}\right)_{\mu\nu}
\end{aligned} \tag{3.3.4}
$$

Similarly we will now find the electronic Hessian. The terms from (3.2.11) needed for this are

$$
\begin{aligned}
E_2(\mathbf{X}) =& Tr\left(\mathbf{f}_a\left[\left[\mathbf{R}_a, \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^-_{\mu\nu}\right]_S, \sum_{\rho>\sigma} X_{\rho\sigma}\mathbf{E}^-_{\rho\sigma}\right]_S\right) \\
&+Tr\left(\left[\mathbf{R}_a, \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^-_{\mu\nu}\right]_S \mathbf{G}\left(\left[\mathbf{R}_a, \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^-_{\mu\nu}\right]_S\right)\right) \\
&+Tr\left(\left[\left[\mathbf{R}_a, \sum_{\mu>\nu} X_{\mu\nu}\mathbf{E}^-_{\mu\nu}\right]_S, \sum_{\rho>\sigma} X_{\rho\sigma}\mathbf{E}^-_{\rho\sigma}\right]_S \mathbf{G}(\mathbf{R}_r)\right)
\end{aligned} \tag{3.3.5}
$$

Differentiating twice with respect to $\mathbf{X}$ and defining a permutation operator $P_{\tau\upsilon,\alpha\beta}$ that permutes the index-pairs $\tau\upsilon$ and $\alpha\beta$, gives

$$
\begin{aligned}
E^{(2)}_{\tau\upsilon\alpha\beta}(\mathbf{X}) = &(1 + P_{\tau\upsilon,\alpha\beta})\, Tr\left(\mathbf{f}_a \left[\left[\mathbf{R}_a, \mathbf{E}^-_{\tau\upsilon}\right]_S, \mathbf{E}^-_{\alpha\beta}\right]_S\right) \\
&+ (1 + P_{\tau\upsilon,\alpha\beta})\, Tr\left(\left[\mathbf{R}_a, \mathbf{E}^-_{\tau\upsilon}\right]_S \mathbf{G}\left(\left[\mathbf{R}_a, \mathbf{E}^-_{\alpha\beta}\right]_S\right)\right) \\
&+ (1 + P_{\tau\upsilon,\alpha\beta})\, Tr\left(\left[\left[\mathbf{R}_a, \mathbf{E}^-_{\tau\upsilon}\right]_S, \mathbf{E}^-_{\alpha\beta}\right]_S \mathbf{G}(\mathbf{R}_r)\right)
\end{aligned} \tag{3.3.6}
$$

3.4 Second-order optimization with Newton's Method

The gradient and Hessian presented in Section 3.3 are explicitly constructed, which is highly expensive step in calculations when attempting to optimize the energy. We therefore need to resort to approximative solutions, here specifically Newton's method, as used by Høst et al.[15] and more closely explained in Helgaker et al.[11]. In this method the energy $Q(\mathbf{X})$ is a Taylor expansion of the energy truncated after second order. $\mathbf{X}$ is, as explained in previous sections, the parameter to be varied in optimization

$$
Q_n(\mathbf{X}) = E^{(0)}_n + \mathbf{X}^T\mathbf{E}^{(1)}_n + \frac{1}{2}\mathbf{X}^T\mathbf{E}^{(2)}_n\mathbf{X} \tag{3.4.1}
$$

The equations to be solved in Newton's method are the linear equations

$$
\mathbf{E}^{(2)}_n\mathbf{X}_n = -\mathbf{E}^{(1)}_n \tag{3.4.2}
$$

where $\mathbf{E}^{(2)}_n$ is the n'th iteration of the electronic Hessian around the expansion point $\mathbf{X} = 0$, $\mathbf{E}^{(1)}_n$ is the corresponding electronic gradient. (3.4.1) is best solved iteratively in a reduced space algorithm, such as the *Davidson algorithm*[9]. We will derive how the Newton equation will look for the embedded system, and start by truncating (3.2.8) to second-order

$$
\begin{aligned}
Q(\mathbf{X}) = &E^{(0)} + 2Tr(\mathbf{f}_a\left[\mathbf{R}_a, \mathbf{X}\right]_S) + Tr(\mathbf{f}_a\left[\left[\mathbf{R}_a, \mathbf{X}\right]_S, \mathbf{X}\right]_S) \\
&+ Tr(\left[\mathbf{R}_a, \mathbf{X}\right]_S \mathbf{G}(\left[\mathbf{R}_a, \mathbf{X}\right]_S)) + 2Tr(\left[\mathbf{R}_a, \mathbf{X}\right]_S \mathbf{G}(\mathbf{R}_r)) \\
&+ Tr(\left[\left[\mathbf{R}_a, \mathbf{X}\right]_S, \mathbf{X}\right]_S \mathbf{G}(\mathbf{R}_r))
\end{aligned} \tag{3.4.3}
$$

We then differentiate each term with respect to $X_{\mu\nu}$ as in the previous section and set $Q'(\mathbf{X}) = 0$. Using that

$$
Tr\left(\left[\mathbf{A}, \left[\mathbf{B}, \mathbf{C}\right]_S\right]_S\right) = Tr\left(\left[\left[\mathbf{A}, \mathbf{B}\right]_S, \mathbf{C}\right]_S\right) \tag{3.4.4}
$$

$$
\left[\mathbf{A}, \left[\mathbf{B}, \mathbf{C}\right]_S\right]_S + \left[\mathbf{B}, \left[\mathbf{C}, \mathbf{A}\right]_S\right]_S + \left[\mathbf{C}, \left[\mathbf{A}, \mathbf{B}\right]_S\right]_S = 0 \tag{3.4.5}
$$

$$\left[\left[\mathbf{R}_a, \mathbf{E}^-_{\mu\nu}\right]_S, \mathbf{X}\right]_S = -\left[\left[\mathbf{E}^-_{\mu\nu}, \mathbf{R}_a\right]_S, \mathbf{X}\right]_S \tag{3.4.6}$$

and (3.3.2), we get

$$\begin{aligned}
4Tr\left(\mathbf{f}_a\left[\mathbf{R}_a, \mathbf{E}_{\mu\nu}\right]_S\right) &+4Tr\left(\left[\mathbf{R}_a, \mathbf{E}_{\mu\nu}\right]_S \mathbf{G}\left(\mathbf{R}_r\right)\right) \\
&= -4Tr\left(\mathbf{f}_a\left[\left[\mathbf{R}_a, \mathbf{X}\right]_S, \mathbf{E}_{\mu\nu}\right]_S\right) \\
&-2Tr\left(\mathbf{f}_a\left[\left[\mathbf{X}, \mathbf{E}_{\mu\nu}\right]_S, \mathbf{R}_a\right]_S\right) \\
&-4Tr\left(\left[\mathbf{R}_a, \mathbf{E}_{\mu\nu}\right]_S \mathbf{G}\left(\left[\mathbf{R}_a, \mathbf{X}\right]_S\right)\right) \\
&-4Tr\left(\left[\left[\mathbf{R}_a, \mathbf{X}\right]_S, \mathbf{E}_{\mu\nu}\right]_S \mathbf{G}\left(\mathbf{R}_r\right)\right) \\
&-2Tr\left(\left[\left[\mathbf{X}, \mathbf{E}_{\mu\nu}\right]_S, \mathbf{R}_a\right]_S \mathbf{G}\left(\mathbf{R}_r\right)\right)
\end{aligned} \tag{3.4.7}$$

The next step is to expand the commutators containing $\mathbf{E}_{\mu\nu}$. We also need to apply the cyclic trace-rule, (3.3.3) and set the equation for all μ, ν. We then introduce the following notation (symmetric and antisymmetric, respectively)

$$\left[\mathbf{M}\right]^S = \frac{1}{2}\left(\mathbf{M} + \mathbf{M}^T\right) \tag{3.4.8}$$

$$\left[\mathbf{M}\right]^A = \frac{1}{2}\left(\mathbf{M} - \mathbf{M}^T\right) \tag{3.4.9}$$

We are now left with

$$\begin{aligned}
&8\left[\mathbf{f}_a\mathbf{R}_a\mathbf{S}\right]^A 8\left[\mathbf{G}\left(\mathbf{R}_r\right)\mathbf{R}_a\mathbf{S}\right]^A \\
&\quad = -\Big[16\mathbf{f}_a\left[\mathbf{R}_a\mathbf{S}\mathbf{X}\right]^S\mathbf{S} - 8\left[\mathbf{f}_a\mathbf{R}_a\mathbf{S}\right]^A\mathbf{X}\mathbf{S} + 16\mathbf{G}\left(\left[\mathbf{R}_a\mathbf{S}\mathbf{X}\right]^S\right)\mathbf{R}_a\mathbf{S} \\
&\quad + 16\mathbf{G}\left(\mathbf{R}_r\right)\left[\mathbf{R}_a\mathbf{S}\mathbf{X}\right]^S\mathbf{S} - 8\left[\mathbf{G}\left(\mathbf{R}_r\right)\mathbf{R}_a\mathbf{S}\right]^A\mathbf{X}\mathbf{S}\Big]^A
\end{aligned} \tag{3.4.10}$$

or its scaled version

$$\begin{aligned}
&-\left[\mathbf{f}_a\mathbf{R}_a\mathbf{S}\right]^A - \left[\mathbf{G}\left(\mathbf{R}_r\right)\mathbf{R}_a\mathbf{S}\right]^A \\
&\quad = \Big[2\mathbf{f}_a\left[\mathbf{R}_a\mathbf{S}\mathbf{X}\right]^S\mathbf{S} - \left[\mathbf{f}_a\mathbf{R}_a\mathbf{S}\right]^A\mathbf{X}\mathbf{S} + 2\mathbf{G}\left(\left[\mathbf{R}_a\mathbf{S}\mathbf{X}\right]^S\right)\mathbf{R}_a\mathbf{S} \\
&\quad + 2\mathbf{G}\left(\mathbf{R}_r\right)\left[\mathbf{R}_a\mathbf{S}\mathbf{X}\right]^S\mathbf{S} - \left[\mathbf{G}\left(\mathbf{R}_r\right)\mathbf{R}_a\mathbf{S}\right]^A\mathbf{X}\mathbf{S}\Big]^A
\end{aligned} \tag{3.4.11}$$

We now have the negative electronic gradient on the left side, and the linear transformation of the Hessian on the trial vector $\mathbf{X}$ on the right, as $-\mathbf{G} = \mathbf{H}\mathbf{X}$. As mentioned, this will be solved in a *reduced room algorithm*, meaning that the Hessian is never ex-

plicitly constructed. The Hessian is replaced by an approximate version, and we get quasi-Newton equations that we solve instead. Among such approximative algorithms are the already mentioned davidson algorithm[9] or augmented Roothan Hall method[15].

3.5 Projection of the orbital rotation operator

The object of the transformation introduced in Section 2.6 is to make it possible, through orbital rotations, to reach all other possible determinants, and thereby find the optimized density matrix and molecular energy. The orbital rotation parameters in $\boldsymbol{\kappa}$ mix between all MOs[12], but not all these parameters affect the wave function, or density matrix. These *redundant* parameters can be excluded without doing harm to our parametrization. It should also be noted that including redundant parameters can result in a considerable problems during optimization, such as singularities in the Hessian[11], which should be avoided.

In RHF, rotations among occupied-occupied orbitals and virtual-virtual orbitals are redundant. The only nonredundant rotations are those between occupied and virtual orbitals. In our scaled density matrix approach, these redundancies are removed by projecting our orbital rotation operator $\mathbf{X}$ so that its only non-zero elements are in the occupied-virtual and virtual-occupied blocks. This projection is done by defining the projectors

$$\mathbf{P} = \mathbf{RS} \tag{3.5.1}$$

$$\mathbf{Q} = 1 - \mathbf{RS} \tag{3.5.2}$$

$\mathbf{P}$ and $\mathbf{Q}$ project onto the occupied and virtual MO space, respectively(Helgaker et al.[11]), and we can now use them to properly project $\mathbf{X}$

$$\begin{aligned} \mathbf{X}_{\text{ov}} =& \mathbf{PXQ}^T + \mathbf{QXP}^T \\ =& \mathbf{PXQ} + \mathbf{QXP} \end{aligned} \tag{3.5.3}$$

Our new projected $\mathbf{R}$ is then

$$\mathbf{R}(\mathbf{X}) = \exp(-\mathbf{X}_{\text{ov}}\mathbf{S})\mathbf{R}\exp(\mathbf{S}\mathbf{X}_{\text{ov}}) \tag{3.5.4}$$

An additional projection is also necessary, but for a different reason. The initial scaled density matrices are all idempotent, but as $\mathbf{R}_a$ is updated, it must be ensured that the complete scaled density matrix $\mathbf{R}$ stays idempotent as well. This is ensured by adding

the projection

$$\begin{aligned}\mathbf{X}_p =&(1-\mathbf{R}_r\mathbf{S})\mathbf{X}(1-\mathbf{R}_r\mathbf{S})^T\\ =&(1-\mathbf{R}_r\mathbf{S})\mathbf{X}(1-\mathbf{S}\mathbf{R}_r)\end{aligned} \tag{3.5.5}$$

to our linear transformation. We call this last projection *Qr-projection*. It ensures idempotency by removing terms containing cross-mixing between the active an inactive part in the total density matrix, and a proof of how the idempotency of the total density matrix is preserved is given in Section 3.7.

3.6 Projected Newton equations

Until now, we have been looking at a version of the AO basis that is not necessarily orthonormal. One of the consequences of this, is that the overlap matrix $\mathbf{S}$ is not restricted to the identity matrix. However, there are great advantages to work in an orthonormal basis[15], and for the rest of this section we will be looking at orthonormalized systems, henceforth called the OAO basis. There are several methods for orthonormalizing our initial atomic functions f_i[7], but maybe the most convenient is the *symmetric orthogonalization* or the Löwdin decomposition which has the advantage that the orthonormal set of functions ψ_i closely resembles the initial atomic functions. This is ensured by minimizing the square difference between f_i and ψ_i

$$\sum_i \int |\psi_i - f_i|^2 d\tau \tag{3.6.1}$$

The generation of ψ_i is performed through a unitary transformation with the unitary matrix U

$$\psi_i = \sum_j f_j (U\Lambda^{-\frac{1}{2}}U^{-1})_{ji} \tag{3.6.2}$$

where $\Lambda^{-\frac{1}{2}}$ is a positive diagonal matrix. An obvious advantage of working in an orthonormalized basis is the fact that since $\mathbf{S} = 1$, matrix-multiplications are simplified. Furthermore, the condition number (a measure of how sensitive the system is to small deviations[19]) is greatly reduced[15]. In 2.6, we introduced a number of conditions, the symmetry condition (2.6.1), the trace condition (2.6.2) and the idempotency condition (2.6.3) for the scaled density matrix. In the new basis, the symmetry condition stays the same, and the equivalents of the trace and idempotency conditions are

$$Tr(\mathbf{R}^u) = \frac{1}{2}N \tag{3.6.3}$$

$$(\mathbf{R}^u)^2 = \mathbf{R}^u \tag{3.6.4}$$

where $\mathbf{R}^u$ is a scaled density matrix in the OAO basis. $\mathbf{P}$ and $\mathbf{Q}$ are also equivalently written as

$$\mathbf{P} = \mathbf{R}^u \tag{3.6.5}$$

$$\mathbf{Q} = 1 - \mathbf{R}^u \tag{3.6.6}$$

We are now ready to look at the main object of this section, namely to find the projected Newton equations. In Section 3.5, we saw that the orbital rotation operator $\mathbf{X}$ has to projected in order to avoid redundancies. It is however also necessary to project the equations as a whole in order to obtain the right unitary transformations. Høst et al.[15] found that the PQ-projected Newton equation $-\mathbf{G} = \mathbf{HX}$ for the regular non-embedded system can be written as

$$\begin{aligned} -\left(\mathbf{F}^{\mathrm{ov}} - \mathbf{F}^{\mathrm{vo}}\right) = &\left(\mathbf{F}^{\mathrm{vv}} - \mathbf{F}^{\mathrm{oo}}\right)\mathbf{X}_{\mathrm{ov}} + \mathbf{X}_{\mathrm{ov}}\left(\mathbf{F}^{\mathrm{vv}} - \mathbf{F}^{\mathrm{oo}}\right) \\ &+\mathbf{G}^{\mathrm{ov}}\left([\mathbf{R}, \mathbf{X}_{\mathrm{ov}}]\right) - \mathbf{G}^{\mathrm{vo}}\left([\mathbf{R}, \mathbf{X}_{\mathrm{ov}}]\right) \end{aligned} \tag{3.6.7}$$

where $\mathbf{F}^{\mathrm{ov}} = \mathbf{PFQ}$ and so on. So how does this translate for the embedded scheme? The equivalent Newton equation is (3.4.11), in the AO basis. We now need to convert this into the OAO basis, substitute for the projected orbital rotation operator $\mathbf{X}_{\mathrm{ov}}$ and then PQ-project the complete result. This gives

$$\begin{aligned} &-\frac{1}{2}\left[-\mathbf{F}^{\mathrm{ov}} + \mathbf{F}^{\mathrm{vo}} + \mathbf{G}^{\mathrm{vo}}(\mathbf{R}_r^u) - \mathbf{G}^{\mathrm{ov}}(\mathbf{R}_r^u)\right] \\ &= -\frac{1}{2}\left[\left(\mathbf{F}^{\mathrm{vv}} - \mathbf{F}^{\mathrm{oo}}\right)\mathbf{X}_{\mathrm{ov}} + \mathbf{X}_{\mathrm{ov}}\left(\mathbf{F}^{\mathrm{vv}} - \mathbf{F}^{\mathrm{oo}}\right)\right. \\ &- \mathbf{G}^{\mathrm{vo}}\left([\mathbf{R}_a^u, \mathbf{X}_{\mathrm{ov}}]\right) + \mathbf{G}^{\mathrm{ov}}\left([\mathbf{R}_a^u, \mathbf{X}_{\mathrm{ov}}]\right) \\ &\left. - \left(\mathbf{G}^{\mathrm{vv}}(\mathbf{R}_r^u) - \mathbf{G}^{\mathrm{oo}}(\mathbf{R}_r^u)\right)\mathbf{X}_{\mathrm{ov}} + \mathbf{X}_{\mathrm{ov}}\left(\mathbf{G}^{\mathrm{vv}}(\mathbf{R}_r^u) - \mathbf{G}^{\mathrm{oo}}(\mathbf{R}_r^u)\right)\right] \end{aligned} \tag{3.6.8}$$

where $\mathbf{R}_a^u$ and $\mathbf{R}_r^u$ are the equivalents of $\mathbf{R}_a$ and $\mathbf{R}_r$ in the OAO basis. This is scaled by a factor $-\frac{1}{2}$ in comparison to the original equation(this stems from how (3.4.8) and (3.4.9) are defined). This is not a problem, it should just be remembered during implementation.

As stated i Section 3.5, for our scheme, it is not sufficient to remove the occupied-occupied and virtual-virtual redundancies. In order to preserve idempotency for the complete system, we also need to implement Qr-projection. Unlike the PQ-projection shown above, this does not produce simplified equations, so the projections will not be

attempted to be written out here. It should just be remembered during implementation that both $\mathbf{X}$ and the complete equations should be projected with $Qr^u = 1 - \mathbf{R}_r^u$.

3.7 Idempotency

As mentioned in Section 2.6 an important trait of the scaled density matrices is that they are idempotent

$$\mathbf{RSR} = \mathbf{R} \tag{3.7.1}$$

We construct the initial density matrices $\mathbf{R}$, $\mathbf{R}_a$ and $\mathbf{R}_r$ to be idempotent, but we need to be sure that this is not only true initially, but also after each optimization iteration done on $\mathbf{R}_a$

$$\mathbf{R}^{n+1}\mathbf{S}\mathbf{R}^{n+1} = \mathbf{R}^{n+1} \tag{3.7.2}$$

where

$$\mathbf{R}^{n+1} = \mathbf{R}_a^{n+1} + \mathbf{R}_r \tag{3.7.3}$$

From the structure of $\mathbf{R}_a$ and $\mathbf{R}_r$ we know that initially,

$$\mathbf{R}_a\mathbf{S}\mathbf{R}_r = 0 \tag{3.7.4}$$

The unitary transformations used to update $\mathbf{R}_a$ are

$$\mathbf{R}_a^{n+1} = \exp(-\mathbf{X}_p\mathbf{S})\mathbf{R}_a^n \exp(\mathbf{S}\mathbf{X}_p) \tag{3.7.5}$$

where $\mathbf{X}_p$ here is the Q_r-projected version of the antisymmetric matrix $\mathbf{X}$ for parametrization

$$\begin{aligned}\mathbf{X}_p =& (1 - \mathbf{R}_r\mathbf{S})\mathbf{X}(1 - \mathbf{R}_r\mathbf{S})^T \\ =& (1 - \mathbf{R}_r\mathbf{S})\mathbf{X}(1 - \mathbf{S}\mathbf{R}_r)\end{aligned} \tag{3.7.6}$$

with the following expansion

$$\exp(\mathbf{S}\mathbf{X}_p) = 1 + \mathbf{S}\mathbf{X}_p + \frac{1}{2}(\mathbf{S}\mathbf{X}_p)^2 + O(\mathbf{X}_p^3) \tag{3.7.7}$$

When we now want to prove (3.7.2), this is equivalent to

$$\begin{aligned}\mathbf{R}^{n+1}\mathbf{S}\mathbf{R}^{n+1} =&(\mathbf{R}_a^{n+1}+\mathbf{R}_r)\mathbf{S}(\mathbf{R}_a^{n+1}+\mathbf{R}_r)\\ =&\mathbf{R}_a^{n+1}\mathbf{S}\mathbf{R}_a^{n+1}+\mathbf{R}_r\mathbf{S}\mathbf{R}_a^{n+1}+\mathbf{R}_a^{n+1}\mathbf{S}\mathbf{R}_r+\mathbf{R}_r\mathbf{S}\mathbf{R}_r\\ =&\mathbf{R}_a^{n+1}+\mathbf{R}_r=\mathbf{R}^{n+1}\end{aligned} \tag{3.7.8}$$

From this, we see that proving (3.7.2) is equivalent to proving that

$$\mathbf{R}_a^{n+1}\mathbf{S}\mathbf{R}_r=\mathbf{R}_r\mathbf{S}\mathbf{R}_a^{n+1}=0 \tag{3.7.9}$$

Since $\mathbf{R}_a^{n+1}\mathbf{S}\mathbf{R}_r$ and $\mathbf{R}_r\mathbf{S}\mathbf{R}_a^{n+1}$ are the transpose of each other, proving that one of them is equal to zero is sufficient. Insert (3.7.5) into (3.7.9), and using the expansion (3.7.7), gives

$$\begin{aligned}\mathbf{R}_a^{n+1}\mathbf{S}\mathbf{R}_r =&\exp(-\mathbf{X}_p\mathbf{S})\mathbf{R}_a^n\exp(\mathbf{S}\mathbf{X}_p)\mathbf{S}\mathbf{R}_r\\ =&\exp(-\mathbf{X}_p\mathbf{S})\mathbf{R}_a^n\left(1+\mathbf{S}\mathbf{X}_p+\frac{1}{2}\mathbf{S}\mathbf{X}_p\mathbf{S}\mathbf{X}_p+O(\mathbf{X}_p^3)\right)\mathbf{S}\mathbf{R}_r\end{aligned} \tag{3.7.10}$$

The first term, we already know from (3.7.4) is zero. The rest of the terms all end with $\mathbf{X}_p\mathbf{S}\mathbf{R}_r$ which can be proved to be zero as well

$$\begin{aligned}\mathbf{X}_p\mathbf{S}\mathbf{R}_r =&\left(1-\mathbf{R}_r\mathbf{S}\right)\mathbf{X}\left(1-\mathbf{S}\mathbf{R}_r\right)\mathbf{S}\mathbf{R}_r\\ =&\left(1-\mathbf{R}_r\mathbf{S}\right)\mathbf{X}(\mathbf{S}\mathbf{R}_r-\mathbf{S}\mathbf{R}_r\mathbf{S}\mathbf{R}_r)\end{aligned} \tag{3.7.11}$$

since $\mathbf{R}_r$ is idempotent ($\mathbf{S}\mathbf{R}_r=\mathbf{S}\mathbf{R}_r\mathbf{S}\mathbf{R}_r$). We have now proved that $\mathbf{R}_a^{n+1}\mathbf{S}\mathbf{R}_r=0$ and thereby (3.7.2). This means that if $\mathbf{R}_r$ and $\mathbf{R}_a^{n+1}$ are idempotent, so is $\mathbf{R}^{n+1}$.

One might note that even though this is true theoretically, if one is not careful with what numerical methods used to update $\mathbf{R}_a$, idempotency might not be preserved. If this is the case, the method should be changed so that idempotency is in fact preserved, or purification should be implemented as a step in each iteration.

Chapter 4

Practical considerations surrounding the implementation

In the previous chapter theoretical consequences of partitioning of the density matrix were discussed. However, since we are not only interested in the theory, but also in being able to run calculations using our new method, we will here briefly introduce a few practicalities concerning how to perform these calculations. All calculations have been run in the density-based framework of the LSDalton[1] package.

4.1 Start guess

As stated by Lenthe et al.[29], the quality of the start guess for the orbital densities is of high importance for any type of calculation. Starting with an inaccurate guess can force the SCF calculation to run through so many iterations that the procedure ends up being exceedingly slow. In some instances convergence might not be reached at all. Especially for the embedded scheme a bad start guess can have damaging results as $\mathbf{R}_r$ is not updated in the optimization and therefore will remain the same no matter how many iterations are performed.

Density matrix from atomic input

Lenthe et al.[29] proposed a procedure where the initial density is found from the atoms and constructs a density matrix that is a *superposition of the atomic densities* (SAD). These are block-diagonal, but nonidempotent. Several schemes exist to solve this problem, maybe the most intuitively obvious one, and the one currently applied in our calculations, is to from the initial density matrix create the Fock matrix which is then diagonalized. From this diagonalized Fock matrix, we find our molecular coefficients, and thereby our idempotent density matrix (in the AO basis). The problem with this method is that it requires diagonalization of the complete Fock matrix, which is a costly affair. An

alternative that avoids this impracticality is to purify the density matrix[11;22]. This is an iterative fixed-point procedure[19] which has been shown to give fast convergence, and can be applied to even the most nonidempotent density matrices.

An alternative to the SAD start guess is the *extended Hückel method*, where the Hartree-Fock equations (2.4.6) are solved in an approximate manner to find the wave function[14]. The diagonal elements on the Hamiltonian H_{ii} are chosen from table values, and the off-diagonal elements H_{ij} are found as a scaled average of H_{ii} and H_{jj} multiplied with the non-zero overlap matrix. This method has been proven to give good geometry optimization, but is not necessarily so good for other calculations. This can be somewhat helped by making the Hamiltonian elements dependent on charges[8], with the disadvantage that the method now has to be iterative[25]. In our scheme however, this is not a problem as we are already dealing with a self-consistent method, and the Hückel method would only be used to make an initial guess of the density matrix.

Partitioning into active- and rest densities

In LSDalton[1] there are already procedures in place for acquiring a good SAD start guess from the input molecular geometry, but in the multi-level scheme, we have to make sure that the density matrix is partitioned in such a way that we retain the necessary properties in both the active and rest matrices, as well as in the total(see (2.6.1),(2.6.2) and (2.6.3)).

We have implemented two schemes for this in LSDalton. The first is a procedure where the molecular coefficients obtained through diagonalization are partitioned, and the active- and rest densities are found from these. Both the total and the partitioned densities produced are idempotent.

The disadvantage of the procedure described above is the fact that it requires the coefficient matrices that can only be found through diagonalization of the Fock matrix, which we have already discussed is costly. A commonly used procedure which instead allows direct partition the total density matrix is Cholesky decomposition[3;23]. This method is used in among others methods Multi level CC[23], and the active and rest densities are selected to be localized on the active and rest atoms, respectively. The method produces orthonormal, localized orbitals[3]. The diagonal elements over a specific threshold (typically 0.2[23]) on active atoms are selected and the corresponding column and row are removed from the total density matrix. This procedure is repeated until all the active elements have been removed, and we are then left with the rest density. This partitioning makes sure the wave function is consistently defined throughout the entire system.

4.2 Level-shift

As explained in Section 3.4, we solve our system by Taylor-expanding the energy as a function of $\mathbf{X}$ around the expansion point $\mathbf{X} = 0$ and truncating after second order as in eq. (3.4.1). For convenience we repeat it here

$$Q(\mathbf{X}) = E^{(0)} + \mathbf{X}^T\mathbf{E}^{(1)} + \frac{1}{2}\mathbf{X}^T\mathbf{E}^{(2)}\mathbf{X} \tag{4.2.1}$$

This can also be described as a trust-region method, or a quadratic model[18;16]. When used correctly, this method should produce quadratic convergence.

The trust-region is defined as the region where (4.2.1) is a good approximation to the non-truncated expression. The second-order expression gives us a hyberbola with h as the radius. We define $\mathbf{X}$ to be within the trust-region if $\|\mathbf{X}\| \leq h$ where $\|\mathbf{X}\|$ is the Frobenius norm[18]. If this is true and the Hessian is positive definite, we can iteratively solve the simple equation

$$\mathbf{E}^{(2)}\mathbf{X} = -\mathbf{E}^{(1)} \tag{4.2.2}$$

However, if the conditions above are not upheld, $\mathbf{X}$ must be constrained in such a way that we do not go beyond the trust-region. This we do by introducing a Lagrangian

$$L(\mathbf{X}, \mu) = E^{(0)} + \mathbf{X}^T\mathbf{E}^{(1)} + \frac{1}{2}\mathbf{X}^T\mathbf{E}^{(2)}\mathbf{X} - \frac{1}{2}\mu(\mathbf{X}^T\mathbf{X} - h^2) \tag{4.2.3}$$

with μ the Lagrangian multiplier to ensure that we stay inside the trust-region. We call μ the level-shift. The equation now to be solved to obtain minimization is

$$(\mathbf{E}^{(2)} - \mu\mathbf{I})\mathbf{X} = -\mathbf{E}^{(1)} \tag{4.2.4}$$

The problem now is to determine μ so that the stationary point found is in fact a minimum, as there are many possible solutions to (4.2.4). In the conjugate residual with optimal trial vectors (CROP) algoritm[30] implemented in the LSDalton package[1], the level shift is determined from the HOMO-LUMO gap, which is problematic as this property really does not make sense in our system and will cause extremely slow convergence. A better approach is to use the line-search procedure, which will not be further discussed in this thesis, but refer to Høyvik et al.[16] for a detailed description.

Chapter 5

Results

In order to test the developed procedure, various test runs were performed. The calculations were performed using the LSDalton[1] package. We have used Cholesky decomposition for partitioning.

Table 5.1 shows computed energy values at completed SCF for a system containing two water molecules where one is active and the other inactive. Calculations have been performed with four different basis sets.

Table 5.1: Energy of two water molecules [a.u.]

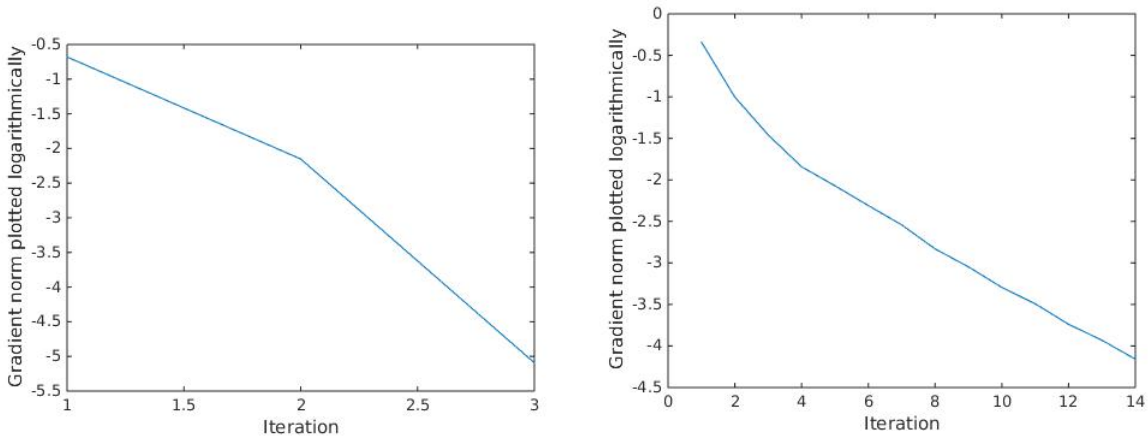
Basis set	Standard HF	Multi-level HF	Difference
STO-3G	-149.934809	-149.903011	-0.031798
6-31G	-151.979938	-151.922755	-0.057183
cc-pVDZ	-152.062705	-152.029779	-0.032926
cc-pVTZ	-152.121240	-152.088265	-0.032975

Similarly, Table 5.1 shows computed energy values at completed SCF for a system of four water molecules, where one water molecule is active and the remaining three are inactive. Calculations have been performed for three different basis sets.

Table 5.2: Energy of four water molecules [a.u]

Basis set	Standard HF	Multi-level HF	Difference
STO-3G	-293.235140	-293.209950	-0.025190
6-31G	-298.309056	-298.137738	-0.171318
cc-pVDZ	-298.509631	-298.311472	-0.198159

The gradient norms of the two-water system in the STO-3G basis set and the four-water system in the cc-pVDZ basis set have been plotted logarithmically against the of iteration number in Figure 5.1. From the figure we see that the simple system indicates quadratic convergence, while the more extensive system shows no such trend.



(a) Two water molecules, STO-3G

(b) Four water molecules, cc-pVDZ

Figure 5.1: Gradient norms for two different systems are plotted logarithmically against iteration number. System 5.1a contains two water molecules where one is active, one inactive, with basis set STO-3G. System 5.1b contains four water molecules where one is active, three inactive, with basis set cc-pVDZ

A slightly more extensive system considered is that of a pyrimidine molecule surrounded by nine water molecules, as shown in Figure 5.2. The geometry has been roughly optimized in Avogadro[10], and we choose the pyrimidine molecule to be active, and the water molecules to be inactive.

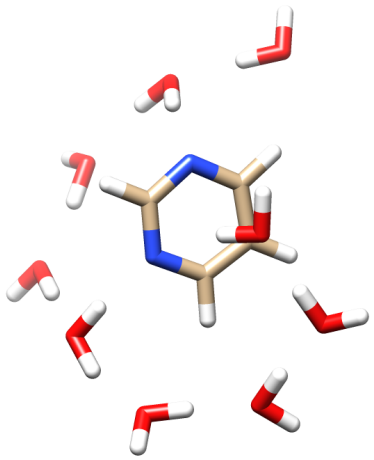


Figure 5.2: Pyrimidine molecule surrounded by nine water molecules

Calculations were run on the pyrimidine system using the basis sets STO-3G and 6-31G. Table 5.3 shows the final results of the calculations at completed SCF for the two different basis sets.

Table 5.3: Energy of pyrimidine molecule surrounded by nine water molecules [a.u.]

Basis set	Normal	Multi-level	Difference
STO-3G	-899.517396	-899.460102	-0.057294
6-31G	-916.498590	-915.654250	-0.844340

Chapter 6

Discussion

This thesis involves the development and implementation of equations fit for the multi-level system. This has been done for both the evaluation of the energy and for the solution of the Newton equations. For the latter the equations have been proven correct through numerical differentiation and the implementation is verified through the fact that we get quadratic convergence close to the minimum for simple systems that do not require level-shift. Cholesky decomposition has been implemented to ensure correct partitioning, and additional projections have been added to maintain idempotency of the total system throughout the optimization process.

In the previous chapter, we saw that all the MLHF energy values are comparable to, but always higher than, those computed from the normal total-density scheme. This follows directly from the variational theorem[5]. The objective of this thesis was to reduce computational cost, but we do not as yet have any proof that this is in fact the case. This is because, although up and running, our system is not yet satisfactorially implemented. In order to verify (or not) that the MLHF method does in fact lower computation costs, we need to solve this problem. The following discussion is about steps we might take in order to make the scheme a real alternative to other procedures.

First of all, the start guess is crucial as we keep $\mathbf{R}_r$ constant. As mentioned in Chapter 4, we should consider other methods than SAD[29], e.g the extended Hückel method[14;25]. The next point is that construction of the **HX** linear transformations is a costly affair, especially if we are dealing with large systems, as is the aim of this thesis. In our calculations we have used an algorithm that requires this construction as in eq. (3.6.8). This gives us accurate results, but is quite costly because we are forced to repeatedly calculate two-electron integrals. For standard HF, methods have been developed to get around this, among them the Roothan-Hall(RH) and augmented Roothan-Hall(ARH) methods[15]. In the first method the two-electron contributions are completely ignored, which will greatly compromise the accuracy of our results. In the second, however, quite good results are obtained by constructing an approximative Hessian from approximative two-electron integrals. In further work, it should be strongly considered whether to

implement an analogous scheme fit to the the multi-level scheme in order to reduce calculation costs.

When solving the linear transformations preconditioning is required. This means to alter the linear transformations in such a way that we lower the condition number and therefore simplify calculations. In LSDalton[1] this method is already in place, but not in a way fit for our scheme. This must be resolved before computations can be run using MLHF on big systems.

Another issue concerning the solution of the linear transformations is the use of level-shift[18]. As mentioned in Chapter 4, algorithms for choosing an appropriate level-shift needs to be in place in order to obtain quadratic convergence. There are in LSDalton[1] available routines for performing a line-search algorithm, a next step would be to properly link these for application in our calculations. The default in the ARH scheme cannot be used because of its dependence on the HOMO-LUMO gap.

Also note that in our implementation of Cholesky decomposition in LSDalton[1], it is a requirement that in the initial total density matrix, the elements of the active space are the ones first on the diagonal. An improved version of this routine should be able to partition the total density regardless of the arrangement of atoms in the input file.

Before we conclude this discussion it should be mentioned that in this thesis we have only described a scheme where $\mathbf{R}_r$ is kept constant. It might be considered whether it is desirable to instead optimize this part as well, but with a less accurate method than for $\mathbf{R}_a$, as in MLCC[23]. Another possibility is to combine MLHF and MLCC in such a way that the total system is partitioned and treated as in MLHF, but the active part is further partitioned and treated with MLCC. Other things to consider would be whether to expand the scheme to contain more that two levels or if it would be convenient to work in the MO basis as opposed to the AO basis.

Chapter 7

Conclusion

The multi-level scheme presented in this thesis is shown to give comparable, but higher results from energy calculations than the non-embedded scheme. This is as expected as only parts of the density matrix is optimized, and we therefore are dealing with a more crude system than in standard Hartree-Fock calculations. The work done in this thesis involves the development and implementation of equations fit for the multi-level scheme. This has been done for both the evaluation of the energy and the solution of the Newton equations. For the latter the equations have been proven correct through numerical differentiation and the implementation is verified through the fact that we get quadratic convergence close to the minimum for simple systems that do not require level-shift. Cholesky decomposition has been implemented to ensure correct partitioning, and additional projections have been added to maintain idempotency of the total system throughout the optimization process.

However, as the procedure is not yet completely developed, we cannot as yet make any conclusions as to whether computation costs are remedied. A number of improvements have been proposed and it is to be hoped that once these are implemented, the procedure might become an advantageous one.

Chapter 8

Future work

As repeatedly stated, a number of alterations are necessary to make the MLHF scheme an advantageous method. Firstly, a preconditioning of the linear transformations suited for our system is needed. Secondly, a method for creating a good level-shift must be implemented, for instance the line-search procedure. Thirdly, the method should be considered being made suitable for approximative solution of the Newton equations with methods such as ARH. Lastly, we do not yet know the full effect the choice of start guess has on our results. The Hückel method and other methods for obtaining initial density matrices should therefore be implemented in order to compare run-times and so on with the current SAD start guess.

Bibliography

[1] K. Aidas, C. Angeli, K.L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E.K. Dalskov, U. Ekström, T. Enevoldsen, J.J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A.C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.M. Høyvik, M.F. Iozzi, B. Jansík, H. J.Aa. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O.B. Lutnæs, J.I. Melo, K.V. Mikkelsen, R.H. Myhre, C. Neiss, C.B. Nielsen, P. Norman, J. Olsen, J.M.H. Olsen, A. Osted, M.J. Packer, F. Pawlowski, T.B. Pedersen, P.F. Provasi, S. Reine, Z. Rinkevicius, T.A. Ruden, K. Ruud, V.V. Rybkin, P. Sałek, C.C.M. Samson, A.S. de Merás, T. Saue, S.P.A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, and H. Ågren. The dalton quantum chemistry program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(3):269–284, 2014.

[2] R. A. Alberty and R. J. Silbey. *Physical Chemistry, l st Ed.* John Wiley Sons Inc., New York, 1992.

[3] F. Aquilante, T.B. Pedersen, A.S. de Merás, and H. Koch. Fast noniterative orbital localization for large molecules. *The Journal of chemical physics*, 125(17):174101, 2006.

[4] D. P. Atkins. *Atkins' Physical Chemistry.* Oxford university press, 2010.

[5] P. W. Atkins and R. S. Friedman. *Molecular quantum mechanics.* Oxford university press, 2011.

[6] L. E. Ballentine. *Quantum mechanics: a modern development, 2nd Ed.* World scientific, 2015.

[7] B.C. Carlson and J.M. Keller. Orthogonalization procedures and the localization of wannier functions. *Physical Review*, 105(1):102, 1957.

[8] L. C. Cusachs and James W. Reynolds. Selection of molecular matrix elements from atomic data. *The Journal of Chemical Physics*, 43(10):S160–S164, 1965.

[9] E. R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *Journal of Computational Physics*, 17(1):87–94, 1975.

[10] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics*, 4(1):17, 2012.

[11] T. Helgaker, P. Jorgensen, and J. Olsen. *Molecular electronic-structure theory*. John Wiley & Sons, 2014.

[12] T. Helgaker, H. Larsen, J. Olsen, and P. Jørgensen. Direct optimization of the ao density matrix in hartree–fock and kohn–sham theories. *Chemical Physics Letters*, 327(5):397–403, 2000.

[13] S. Höfener, A. S. P. Gomes, and L. Visscher. Molecular properties via a subsystem density functional theory formulation: A common framework for electronic embedding. *The Journal of chemical physics*, 136(4):044104, 2012.

[14] R. Hoffmann. An extended hückel theory. i. hydrocarbons. *The Journal of Chemical Physics*, 39(6):1397–1412, 1963.

[15] S. Høst, J. Olsen, B. Jansík, L. Thøgersen, P. Jørgensen, and T. Helgaker. The augmented roothaan–hall method for optimizing hartree–fock and kohn–sham density matrices. *The Journal of chemical physics*, 129(12):124106, 2008.

[16] I.M. Høyvik, B. Jansik, and P. Jørgensen. Trust region minimization of orbital localization functions. *Journal of chemical theory and computation*, 8(9):3137–3146, 2012.

[17] S. Humbel, S. Sieber, and K. Morokuma. The imomo method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and sn2 reaction: Rcl+ cl-. *The Journal of chemical physics*, 105(5):1959–1967, 1996.

[18] I.M. Høyvik. Local hartree-fock orbitals, 2013.

[19] E. Kreyszig. *Advanced engineering mathematics*. John Wiley & Sons, 2010.

[20] A.R. Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.

[21] I. N. Levine and P. H. I. Learning. *Quantum chemistry*, volume 6. Pearson Prentice Hall Upper Saddle River, NJ, 2009.

[22] R. McWeeny. Some recent advances in density matrix theory. *Rev. Mod. Phys.*, 32:335–369, Apr 1960.

[23] R. H. Myhre, S. de Merás, and H. Koch. Multi-level coupled cluster theory. *The Journal of chemical physics*, 141(22):224105, 2014.

[24] R. H. Myhre and S. de Merás. The extended cc2 model ecc2. 111:1109–1118, Jul 2013.

[25] R. Rein, N. Fukuda, H. Win, G. A. Clarke, and F. E. Harris. Iterative extended huckel theory. *The Journal of Chemical Physics*, 45:4743–4744, 1966.

[26] M. A. San Miguel, C. J. Calzado, and J. F. Sanz. Molecular dynamics simulations of na deposition on the tio 2 (110) surface. *Surface science*, 409(1):92–100, 1998.

[27] R. Shankar. *Principles of quantum mechanics*. Plenum Press, 1980.

[28] A. Szabo and N. S. Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.

[29] J. H. Van Lenthe, R. Zwaans, H. J. J. Van Dam, and M. F. Guest. Starting scf calculations by superposition of atomic densities. *Journal of computational chemistry*, 27(8):926–932, 2006.

[30] M. Ziółkowski, V. Weijo, P. Jørgensen, and J. Olsen. An efficient algorithm for solving nonlinear equations with a minimal number of trial vectors: Applications to atomic-orbital based coupled-cluster theory. *The Journal of chemical physics*, 128(20):204105, 2008.