SCF algorithms for Hartree-Fock electronic calculations*

Eric Cancès
CERMICS, Ecole Nationale des Ponts et Chaussées,
6 & 8 avenue Blaise Pascal, Cité Descartes,
F-77455 Marne-la-Vallée, France.
cances@cermics.enpc.fr

November 26, 1999

Abstract

This paper presents some mathematical results on SCF algorithms for solving the Hartree-Fock problem. In the first part of the article the focus is on two classical SCF procedures, namely the Roothaan algorithm and the level-shifting algorithm. It is demonstrated that the Roothaan algorithm either converges towards a solution to the Hartree-Fock equations or oscillates between two states which are not solution to the Hartree-Fock equations, any other behavior (oscillations between more than two states, "chaotic" behavior, ...) being excluded. The level-shifting algorithm is then proved to converge for large enough shift parameter, whatever the initial guess. The second part of the article details the convergence properties of a new algorithm recently introduced by Le Bris and the author, the so-called Optimal Damping Algorithm (ODA). Basic numerical simulations pointing out the principal features of the various algorithms under study are also provided.

1 Introduction

The Hartree-Fock (HF) model is a standard tool for computing an approximation of the ground state of a molecular system within the Born-Oppenheimer setting. From a mathematical viewpoint, the HF model gives rise to a nonquadratic constrained minimization problem for the numerical solution of which iterative procedures are needed; such procedures are referred to as Self-Consistent Field (SCF) algorithms. The solution to the HF problem can be obtained either by directly minimizing the HF energy functional [7, 12, 18, 26] or by solving the associated Euler-Lagrange equations, the so-called Hartree-Fock equations [21, 22, 23].

SCF algorithms for solving the HF equations are in general much more efficient than direct energy minimization techniques. However, these algorithms do not *a priori* ensure the decrease of the energy and they may lead to convergence problems [24]. For instance, the famous Roothaan algorithm (*see* [22] and section 4) is known to sometimes lead to stable oscillations between two states, none of them being a solution to the HF problem. This situation may occur even for simple chemical systems (see section 4).

Many articles have been devoted to the important issue of the SCF convergence. The behavior of the Roothaan algorithm is notably investigated in [2, 13] and in

[*] To appear in *Mathematical models and methods for ab initio Quantum Chemistry*, M. Defranceschi and C. Le Bris (Eds.), in preparation for Lecture Notes in Chemistry, Springer.

[27, 28]. In [2, 13] convergence difficulties are demonstrated for elementary two-dimensional models; in [27, 28], a stability condition of the Roothaan algorithm in the neighbourhood of a minimum of the HF energy is given for closed-shell systems. More sophisticated SCF algorithms for solving the HF equations have also been proposed to improve the convergence using various techniques like for instance damping [11, 29] or level-shifting [23]. Damping (as implemented in [29]) cures some convergence problems but many other remain. Numerical tests confirm that the level-shifting algorithm converges towards a solution to the HF equations for large enough shift parameters; a perturbation argument is provided in [23] to prove this convergence in the neighborhood of a stationary point. Unfortunately there is no guarantee that the so-obtained critical point of the HF energy functional is actually a minimum (even local); in addition, the level-shifting algorithm is known to only offer a slow speed of convergence. In practice, the most commonly used SCF algorithm is at the present time the Direct Inversion in the Iteration Space (DIIS) algorithm [21]. Numerical tests show that this algorithm is very efficient in most cases, but that it sometimes fails.

The present article belongs to a series of articles [3, 4, 5] devoted to the SCF algorithms.

Our first purpose here is to report on recent mathematical results on the convergence properties of the Roothaan and of the level-shifting algorithms. Section 4 concerns the Roothaan algorithm, which is the most “natural” algorithm for solving the HF equations. Its is demonstrated that the Roothaan algorithm either converges towards a solution to the HF equations or oscillates between two states which are not solution to the HF equations, any other behavior being excluded. This theoretical result is in accordance with the numerical experiments. It is then explained in Section 5 why the introduction of a “level-shift” makes the algorithm converge. The mathematical proofs are presented in the context of the *finite dimension* approximations of the HF problem obtained by a Galerkin method with a finite basis of atomic orbitals or plane waves, typically. They are consequently much simpler from a technical viewpoint than the proofs detailed in [3] which concern the original *infinite dimension* HF problem.

Recently, new SCF algorithms has been introduced in [4] by Le Bris and the author. They seem to exhibit good convergence properties at least for the chemical systems computed so far. These algorithms have been called Relaxed Contraints Algorithms (RCA) for they can be interpretated as direct minimization procedure of the HF energy which do not care about satisfying at each iteration the nonlinear constraints $D^2 = D$ that characterize admissible density matrices. The second purpose of this article (section 6) is to detail the mathematical proof of the convergence of the basic RCA, namely the Optimal Damping Algorithm (ODA). Section 6 also contains some comments on the connexions between RCA and other algorithms like the level-shifting and the DIIS algorithms.

Before coming up to our main topic, we devote section 2 to a brief presentation of the HF model for readers (especially mathematicians) who are not familiar with Quantum Chemistry. Section 3 collects various general comments that apply to all the SCF algorithms considered in the sequel.

2 A brief presentation of the Hartree-Fock model

The problem under consideration consists in computing *ab initio*, that is to say without using any empirical parameter, the ground state energy of a molecular system made of M nuclei and N electons. Tackling directly the $M+N$-body Schrödinger equation is today, and will probably remain, out of the scope of brute force numerical methods. Various approximations are therefore to be resorted to.

The first approximation that is common to most models of Quantum Chemistry is the so-called *Born-Oppenheimer approximation*. To make short, it consists in considering the nuclei as classical point particles. The Born-Oppenheimer approximation, which has been mathematically founded by Combes and al. [6], lays on the fact that nuclei are much heavier than electrons. The Born-Oppenheimer approximation is almost always valid in Chemistry (except for instance for studying specifically quantum phenomena involving nuclei as proton transfer by tunnel effect) and is therefore almost always used.

Within the Born-Oppenheimer approximation, the searching for the ground state takes the form of two nested minimization problems:

$$\inf \left\{ W(\bar{x}_1, \cdots, \bar{x}_M), \qquad (\bar{x}_1, \cdots, \bar{x}_M) \in \mathbb{R}^{3M} \right\} \tag{1}$$

with

$$W(\bar{x}_1, \cdots, \bar{x}_M) = E_{el}(\bar{x}_1, \cdots, \bar{x}_M) + \sum_{1 \le k < l \le M} \frac{z_k \, z_l}{|\bar{x}_k - \bar{x}_l|}$$

$$E_{el}(\bar{x}_1, \cdots, \bar{x}_M) = \inf \left\{ \langle \psi, H_{\{\bar{x}_k\}} \psi \rangle, \quad \psi \in \mathcal{H}, \quad \|\psi\| = 1 \right\} \tag{2}$$

$$H_{\{\bar{x}_k\}} = - \sum_{i=1}^{N} \frac{1}{2} \Delta_{x_i} - \sum_{i=1}^{N} \sum_{k=1}^{M} \frac{z_k}{|x_i - \bar{x}_k|} + \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|}$$

$$\mathcal{H} = \bigwedge_{i=1}^{N} L^2(\mathbb{R}^3 \times \{|+\rangle, |-\rangle\}, \mathbb{C})$$

In the above expressions, $\bar{x}_k$ denotes the current position in $\mathbb{R}^3$ of the k-th nucleus and z_k its charge. The N electrons are described by a wave function $\psi(x_1, \sigma_1; \cdots; x_N, \sigma_N)$, where x_i and σ_i are respectively the position in $\mathbb{R}^3$ and the spin coordinate of the i-th electron. Each spin coordinate σ_i can take two values here denoted by $|+\rangle$ (spin up) and $|-\rangle$ (spin down). The wave function ψ is a normalized vector of the fermionic Hilbert space $\mathcal{H}$, and therefore satisfies on the one hand the antisymmetry condition

$$\psi(x_{p(1)}, \sigma_{p(1)}; \cdots; x_{p(N)}, \sigma_{p(N)}) = (-1)^{\epsilon(p)} \psi(x_1, \sigma_1; \cdots; x_N, \sigma_N)$$

for any permutation p of $|[1, N]|$ ($\epsilon(p)$ denoting the signature of p), and on the other hand the normalization condition

$$\sum_{\sigma_1, \cdots, \sigma_N} \int_{\mathbb{R}^{3N}} |\psi(x_1, \sigma_1; \cdots; x_N, \sigma_N)|^2 \, dx_1 \, \cdots \, dx_N = 1.$$

The operator $H_{\{\bar{x}_k\}}$ is the so-called electronic hamiltonian. It acts on $\mathcal{H}$; the $\bar{x}_k$ play the role of parameters. It is made of three terms, the first term accounting for the kinetic energy of the electrons, the second and the third terms accounting for nuclei-electrons and electrons-electrons interactions respectively. All physical quantities are expressed in atomic units [19].

Searching for the ground-state of the molecular system thus consists in minimizing the potential energy $W(\bar{x}_1, \cdots, \bar{x}_M)$ by solving the so-called *geometry optimization problem* (1). From the mathematical point of view, problem (1) is an unconstrained minimization problem of finite dimension. We refer the reader to [20, 25] for an overview of the various numerical methods dedicated to geometry optimization.

The specificity of problem (1) is that the function to be minimized, namely the potential energy W, is itself the result (up to the internuclear repulsion term $\sum z_k z_l/|\bar{x}_k - \bar{x}_l|$) of the minimization problem (2) which is usally referred to as the *electronic problem*. We face this time a constrained minimization problem on the infinite dimension space $\mathcal{H}$.

In the sequel, we focus on the electronic problem, which is rewritten (in order to simplify the notations)

$$\inf\left\{\langle\psi, H\psi\rangle, \quad \psi\in\mathcal{H}, \quad \|\psi\| = 1\right\} \tag{3}$$

with

$$\mathcal{H} = \bigwedge_{i=1}^{N} L^2(\mathbb{R}^3\times\{|+\rangle,|-\rangle\},\mathbb{C})$$

$$H = -\sum_{i=1}^{N}\frac{1}{2}\Delta_{x_i} + \sum_{i=1}^{N} V(x_i) + \sum_{1\le i<j\le N}\frac{1}{|x_i - x_j|}$$

$$V(x) = -\sum_{k=1}^{M}\frac{z_k}{|x-\bar{x}_k|}$$

the $\bar{x}_k$ being now fixed parameters in $\mathbb{R}^3$.

The Hartree-Fock approximation is of variational nature. It consists in restricting the set $\{\psi\in\mathcal{H}, \quad \|\psi\| = 1\}$ on which the energy functional $\langle\psi, H\psi\rangle$ is minimized to the set of the Slater determinants, i.e. to the set of the wave functions ψ of the form

$$\psi = \frac{1}{\sqrt{N!}}\det(\phi_i(x_j,\sigma_j)) \tag{4}$$

where the ϕ_i, which are called *molecular orbitals*, satisfy the orthonormality conditions

$$\sum_\sigma\int_{\mathbb{R}^3}\phi_i(x,\sigma)\phi_j(x,\sigma)^*\,dx = \delta_{ij}.$$

A classical calculation (see [19] for instance) gives for any ψ of the form (4)

$$\langle\psi, H\psi\rangle = \mathcal{E}^{HF}(\{\phi_i\})$$

with

$$\begin{aligned}\mathcal{E}^{HF}(\{\phi_i\}) &= \sum_{i=1}^{N}\frac{1}{2}\int_{\mathbb{R}^3}\sum_\sigma|\nabla\phi_i|^2 + \int_{\mathbb{R}^3}\rho_\Phi\,V\\ &\quad+\frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\frac{\rho_\Phi(x)\,\rho_\Phi(x')}{|x-x'|}\,dx\,dx'\\ &\quad-\frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\sum_{\sigma,\sigma'}\frac{|\tau_\Phi(x,\sigma;x',\sigma')|^2}{|x-x'|}\,dx\,dx',\end{aligned}$$

$$\tau_\Phi(x,\sigma;x',\sigma') = \sum_{i=1}^{N}\phi_i(x,\sigma)\,\phi_i(x',\sigma')^*,$$

$$\rho_\Phi(x) = \sum_{i=1}^{N}\sum_\sigma|\phi_i(x,\sigma)|^2.$$

The HF problem thus reads

$$\inf \left\{ \mathcal{E}^{HF}(\{\phi_i\}), \quad \phi_i \in L^2(\mathbb{R}^3 \times \{|+\rangle, |-\rangle\}, \mathbb{C}), \quad \sum_\sigma \int_{\mathbb{R}^3} \phi_\mathrm{i}(\mathrm{x},\sigma)\phi_\mathrm{j}(\mathrm{x},\sigma)^* \,\mathrm{dx} = \delta_\mathrm{ij} \right\}.$$

The mathematical properties of the HF problem have been studied by Lieb and Simon [15] and by P.-L. Lions [16]. The existence of a HF electronic ground state is guaranteed for positive ions ($Z := \sum_{k=1}^M z_k > N$) and neutral systems ($Z = N$). We are not aware of any general existence result for negative ions (the available existence proofs only work for $N < Z+1$). On the other hand, there is a non-existence results for negative ions such that $N > 2Z + M$ [14] (this inequality holds for instance for the ion H^{2-}). As far as we know, uniqueness (of the density ρ at least) is an open problem, probably of outstanding difficulty.

The last step of the approximation procedure consists in approaching the *infinite dimensional* HF problem by a *finite dimensional* HF problem by means of a Galerkin approximation: the HF energy is minimized over the set of molecular orbitals that can be expanded on a given finite basis $\{\chi_p\}_{1\le p\le n}$:

$$\phi_i = \sum_{k=1}^n C_{ki}\chi_k.$$

Denoting by $S = [S_{kl}]$ with

$$S_{kl} = \sum_\sigma \int_{\mathbb{R}_3} \chi_k^* \chi_l$$

the so-called *overlap* matrix, the contraints $\sum_\sigma \int_{\mathbb{R}^3} \phi_i \phi_j^* = \delta_{ij}$ read

$$\delta_{ij} = \sum_\sigma \int_{\mathbb{R}^3} \phi_i\phi_j^* = \sum_\sigma \int_{\mathbb{R}^3} \left(\sum_{l=1}^n C_{li}\chi_l\right)\left(\sum_{k=1}^n C_{kj}\chi_k\right)^* = \sum_{k=1}^n \sum_{l=1}^n C_{kj}^* S_{kl} C_{li},$$

or in matricial form

$$C^*SC = I_N,$$

where I_N denotes the identity matrix of rank N. In addition,

$$\begin{aligned}
\sum_{i=1}^N \frac{1}{2}\int_{\mathbb{R}^3} |\nabla\phi_i|^2 + \int_{\mathbb{R}^3} \rho_\Phi V &= \sum_{i=1}^N \left(\frac{1}{2}\int_{\mathbb{R}^3} |\nabla\phi_i|^2 + \int_{\mathbb{R}^3} V|\phi_i|^2\right) \\
&= \sum_{i=1}^N \left(\frac{1}{2}\int_{\mathbb{R}^3} \left|\nabla \sum_{k=1}^n C_{ik}\chi_k\right|^2 + \int_{\mathbb{R}^3} V \left|\sum_{k=1}^n C_{ik}\chi_k\right|^2\right) \\
&= \sum_{i=1}^N \sum_{k=1}^n \sum_{l=1}^n h_{kl} C_{li} C_{ki}^* \\
&= \mathrm{Tr}\,(hCC^*)
\end{aligned}$$

where h denotes the matrix of the core hamiltonian $-\frac{1}{2}\Delta + V$ in the basis $\{\chi_k\}$:

$$h_{kl} = \frac{1}{2}\sum_\sigma \int_{\mathbb{R}_3} \nabla\chi_k^* \cdot \nabla\chi_l + \sum_\sigma \int_{\mathbb{R}^3} V\chi_k^*\chi_l.$$

Lastly, denoting by

$$(ij|kl) = \sum_\sigma \sum_{\sigma'} \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\chi_i(x)\chi_j(x)^*\chi_k(x')\chi_l(x')^*}{|x-x'|}\,dx\,dx' \quad \text{and} \quad A_{ijkl} = (ij|kl) - (il|kj),$$

the interelectronic repulsion term reads

$$\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\rho_\Phi(x)\,\rho_\Phi(x')}{|x-x'|}\,dx\,dx' - \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{|\tau_\Phi(x,x')|^2}{|x-x'|}\,dx\,dx' = \sum_{i,j,k,l=1}^{n}\sum_{\alpha,\beta=1}^{N} A_{ijkl}C_{i\alpha}C_{j\alpha}^*C_{k\beta}C_{l\beta}^*.$$

The above expressions incite one to introduce the so-called *density matrix*

$$D = CC^*,$$

which permits to write the HF energy under the compact form

$$E^{HF}(D) = \mathrm{Tr}\,(hD) + \frac{1}{2}\mathrm{Tr}\,(G(D)D),$$

where $G(D)$ denotes the contracted product of the 4-index tensor A by D:

$$G(D)_{ij} = (A : D)_{ij} = \sum_{kl} A_{ijkl}D_{kl}.$$

It is easy to see that the matrices D which read $D = CC^*$ with $C \in \mathcal{M}(n,N)$ and $C^*SC = I_N$ are those which satisfy $\mathrm{Tr}\,(SD) = N$ and $DSD = D$. The so-obtained finite dimension HF problem then reads

$$\inf\left\{E^{HF}(D), \quad D \in \mathcal{M}(n,n), \quad D^* = D, \quad \mathrm{Tr}\,(SD) = N, \quad DSD = D\right\}.$$

For the sake of simplicity, we assume in the sequel that the overlap matrix S equals identity, that is to say that the basis $\{\chi_p\}_{1\le p\le n}$ is orthonormal. The general case is recovered by the transformation rules $D \to S^{1/2}DS^{1/2}$, $h \to S^{-1/2}hS^{-1/2}$, $G(D) \to S^{-1/2}G(S^{1/2}DS^{1/2})S^{-1/2}$. The HF problem then reads:

$$\inf\left\{E^{HF}(D), \quad D \in \mathcal{P}\right\}, \tag{5}$$

with

$$\mathcal{P} = \left\{D \in \mathcal{M}(n,n), \quad D^* = D, \quad \mathrm{Tr}\,D = N, \quad D^2 = D\right\}.$$

The following lemma provides a characterization of the critical points of the HF minimization problem (5).

Lemma 1. *For any D, let us denote by $F(D) = h+G(D)$ the Fock matrix associated with D.*

1. *A density matrix $D \in \mathcal{P}$ is a critical point of the HF problem (5) if and only if*

$$\left\{\begin{array}{l} F(D)C = CE \\ C^*C = I_N \\ D = CC^* \end{array}\right. \tag{6}$$

where $E = Diag(\epsilon_1, \epsilon_2, \cdots, \epsilon_N)$ is a $N \times N$ diagonal matrix collecting N eigenvalues of the linear eigenvalue problem

$$F(D)\cdot\phi = \epsilon\,\phi.$$

and where C is a $n \times N$ matrix containing N orthonormal eigenvectors associated with ϵ_1, ϵ_2, ..., ϵ_N. The condition (6) is equivalent to the condition

$$[F(D), D] = 0,$$

where $[\cdot,\cdot]$ denotes the matrix commutator defined for any A and B in $\mathcal{M}(n,n)$ by $[A,B] = AB - BA$.

2. *For $D \in \mathcal{P}$ being a local minimum of the HF problem, it is necessary that ϵ_1, ϵ_2, ..., ϵ_N are the smallest N eigenvalues of $F(D)$ including multiplicity.*

This result is classical; its proof can be read in any textbook of Quantum Chemistry (see [19] for instance).

Remark. The model described above is the so-called General Hartree-Fock (GHF) model. Most often in practice, Quantum Chemistry calculations are performed with spin constraints models like the Restricted Hartree-Fock (RHF), the Unrestricted Hartree-Fock (UHF) or the Restricted Open-shell Hartree-Fock (ROHF) models. The convergence results stated below can be adapted without difficulties to these models. ♢

3 General remarks on SCF algorithms

Various SCF algorithms are studied in the following three sections. All of them consist in generating a sequence (D_k) defined by

$$\left\{ \begin{array}{l} \widetilde{F}_k C_{k+1} = C_{k+1} E_{k+1} \\ C_{k+1}^* C_{k+1} = I_N \\ D_{k+1} = C_{k+1} C_{k+1}^* \end{array} \right. \tag{7}$$

where $E_{k+1} = \mathrm{Diag}(\epsilon_1^{k+1}, \cdots, \epsilon_N^{k+1})$, $\epsilon_1^{k+1} \leq \epsilon_2^{k+1} \leq \cdots \leq \epsilon_n^{k+1}$ being the eigenvalues of the linear eigenvalue problem

$$\widetilde{F}_k \cdot \phi = \epsilon \, \phi,$$

and where C_{k+1} collects N orthonormal eigenvectors associated with ϵ_1^{k+1}, ϵ_2^{k+1}, ..., ϵ_N^{k+1}. The expression of the current Fock matrix $\widetilde{F}_k$ characterizes the algorithm. We have for instance

- $\widetilde{F}_k = F(D_k)$ for the Roothaan algorithm;
- $\widetilde{F}_k = F(D_k) - bD_k$ where b is a positive constant for the level-shifting algorithm;
- $\widetilde{F}_k = F(\widetilde{D}_k)$ for the ODA, where $\widetilde{D}_k$ is a pseudo-density matrix which satisfies the *relaxed* constraints $\widetilde{D}_k^2 \leq \widetilde{D}_k$ and is defined so that the HF energy $E^{HF}(\widetilde{D}_k)$ decreases at each iteration (see section 6).

The procedure consisting in assembling the matrix $D_{k+1} \in \mathcal{P}$ by populating the N molecular orbitals of lowest energies of the current Fock matrix $\widetilde{F}_k$ is referred to as the *aufbau principle*. It is justified by the results stated in Lemma 1. For the matrix D_{k+1} being defined in a unique way, it suffices that $\epsilon_N^{k+1} < \epsilon_{N+1}^{k+1}$. Degeneracies in the spectrum are in general related to the symmetries of the system: in the cases when the system does not exhibit any symmetry, numerical experiments show that the eigenvalues of $\widetilde{F}_k$ are generically non-degenerate for any k, whereas it may not be the case when the system does exhibit symmetries (consider for instance the spherical symmetry of the hamiltonian in the atomic case). Degeneracies create technical difficulties which complicate the theoretical studies on SCF convergence. For the sake of simplicity, we therefore assume from now on that the *uniform well-posedness* (UWP) property introduced in [3] is satisfied:

UWP property: a SCF algorithm of the form (7) with initial guess D_0 will be said to be uniformly well-posed if there exists some positive constant γ such that

$$\epsilon_{N+1}^{k+1} \geq \epsilon_N^{k+1} + \gamma.$$

The consequences of the UWP assumption which will be useful below have been collected in the following lemma, whose proof is postponed until the end of the present section.

Lemma 2. *Let us consider a SCF algorithm of the form (7) with initial guess D_0 which satisfies the UWP property. Then*

1. *The updated density matrix D_{k+1} is defined in a unique way at each iteration; this matrix can be characterized as the minimizer of the variational problem*
$$\inf \left\{ \mathrm{Tr}\,(\widetilde{F}_k D), \quad D \in \mathcal{P} \right\}.$$
2. *For any $D \in \mathcal{M}(n,n)$ such that $D = D^*$, $\mathrm{Tr}\,(D) = N$ and $D^2 \leq D$,*
$$\mathrm{Tr}\,(\widetilde{F}_k D) \geq \mathrm{Tr}\,(\widetilde{F}_k D_{k+1}) + \frac{\gamma}{2}\|D - D_{k+1}\|^2,$$
$\|\cdot\|$ denoting the Hilbert-Schmidt norm defined for any $A \in \mathcal{M}(n,n)$ by $\|A\| = \mathrm{Tr}\,(AA^*)^{1/2}$.

In the sequel, we denote by arg inf $\mathcal{MP}$ the minimizer of the minimization problem $\mathcal{MP}$. We can therefore write
$$D_{k+1} = \arg\inf \left\{ \mathrm{Tr}\,(\widetilde{F}_k D), \quad D \in \mathcal{P} \right\}.$$

Remark. Let us point out that some convergence results can be obtained without resorting to the UWP assumption. In particular, it turns out that the level-shifting algorithm is automatically UWP as soon as the shift parameter is large enough (see [3] for details). It can also be proved that the ODA numerically converges towards an *aufbau* solution to the HF equations *within the GHF setting* and provided the basis is "large enough". We do not detail here the rather technical proof of this assertion. Let us just mention that it is based on a mathematical result by P.-L. Lions [17] related to finite-temperature HF models. Unfortunately, so far as we know, the arguments used in [17] cannot be extended to the RHF, UHF or ROHF models. ♢

Before turning to the study of SCF algorithms, the notion of convergence has to be made precise. We are in fact not able to prove mathematical convergence results of the form "the sequence (D_k) converges towards a minimizer D of the HF problem (5)" for at least two reasons. First, we are solving the Euler-Lagrange equations associated with the HF minimization problem (5), namely the HF equations (6); even in case of convergence we have no argument to conclude that the so-obtained critical point is actually a minimum (even local) of the HF energy. Second, we have no precise description of the topology of the set of the critical points of (5); this lack of information prevents us from proving the convergence of the whole sequence (D_k) towards a solution D to the HF equations. We can at best obtain that $D_{k+1} - D_k$ goes to zero, and that for "large" k, D_k is "close to" a solution to the HF equations (6) satisfying the *aufbau* principle. For instance, it may happen that the HF problem admits a connected manifold of minima; this phenomenon is observed in particular for open-shell atoms because the spherical symmetry of the problem is broken by the HF approximation (this can be related to a mathematical result by Bach, Lieb, Loss and Solovej [1] stating that "there are no unfilled shell" in the HF ground states). We cannot then discriminate between the case when the sequence (D_k) converges towards a point of the manifold and the case when the sequence (D_k) is attracted by the manifold together with a slow drift parallel to the manifold.

We shall consequently adopt here the following two convergence criteria, which are sufficient in practice. We shall say that a SCF algorithm of the form (7) *numerically converges* towards a solution to the HF equations if the sequence (D_k) satisfies

1. $D_{k+1} - D_k \longrightarrow 0$;
2. $[F(D_k), D_k] \longrightarrow 0$;

and that it *numerically converges* towards an *aufbau* solution to the HF equations if the sequence (D_k) satisfies

1. $D_{k+1} - D_k \longrightarrow 0$;
2. $\mathrm{Tr}\ (F(D_k)D_k) - \inf\{\mathrm{Tr}\ (F(D_k)D), \quad D \in \mathcal{P}\} \longrightarrow 0$.

As all norms are equivalent in finite dimension, we do not need to specify the matrix norm in which the variations are evaluated. Let us remark that the latter convergence criterion is stronger than the former one for

$$(\mathrm{Tr}\ (F(D_k)D_k) - \inf\{\mathrm{Tr}\ (F(D_k)D), \quad D \in \mathcal{P}\} \to 0) \quad \Rightarrow ([F(D_k), D_k] \to 0)\,.$$

Let us conclude this section with the

Proof of Lemma 2. Let us denote by D a current matrix such that $D = D^*$, $\mathrm{Tr}\ (D) = N$, $D^2 \leq D$ and by D_{ij} its coefficients in an orthonormal basis in which $\widetilde{F}_k = \mathrm{Diag}(\epsilon_1^{k+1}, \epsilon_2^{k+1}, \cdots, \epsilon_n^{k+1})$ with $\epsilon_1^{k+1} \leq \epsilon_2^{k+1} \leq \cdots \leq \epsilon_n^{k+1}$. In such a basis $D_{k+1} = \mathrm{Diag}(1, \cdots, 1, 0, \cdots, 0)$. As in addition $\mathrm{Tr}\ (D) = \sum_{i=1}^n D_{ii} = N$, we get first

$$\begin{aligned}
\|D_{k+1} - D\|^2 &= \mathrm{Tr}\ ((D_{k+1} - D) \cdot (D_{k+1} - D)) \\
&= \mathrm{Tr}\ \left(D_{k+1}^2\right) + \mathrm{Tr}\ \left(D^2\right) - 2\,\mathrm{Tr}\ (DD_{k+1}) \\
&\leq \mathrm{Tr}\ (D_{k+1}) + \mathrm{Tr}\ (D) - 2\,\mathrm{Tr}\ (DD_{k+1}) \\
&= 2N - 2\sum_{i=1}^N D_{ii} \\
&= 2\sum_{i=N+1}^n D_{ii}.
\end{aligned}$$

Besides $\mathrm{Tr}\ (\widetilde{F}_k D_{k+1}) = \sum_{i=1}^N \epsilon_i^{k+1}$, $\mathrm{Tr}\ (\widetilde{F}_k D) = \sum_{i=1}^n \epsilon_i^{k+1} D_{ii}$ and

$$0 \leq D_{ii} \leq 1, \qquad \text{for any } 1 \leq i \leq n$$

for $D^2 \leq D = D^*$ implies $|D_{ii}|^2 + \sum_{j \neq i} |D_{ij}|^2 \leq D_{ii}$. Putting together the above results, we obtain

$$\begin{aligned}
\mathrm{Tr}\ (\widetilde{F}_k D) &= \sum_{i=1}^n \epsilon_i^{k+1} D_{ii} \\
&\geq \sum_{i=1}^N \epsilon_i^{k+1} D_{ii} + \sum_{i=N+1}^n (\epsilon_N^{k+1} + \gamma) D_{ii} \\
&= \sum_{i=1}^N \epsilon_i^{k+1} D_{ii} + \epsilon_N^{k+1} \sum_{i=N+1}^n D_{ii} + \gamma \sum_{i=N+1}^n D_{ii} \\
&= \sum_{i=1}^N \epsilon_i^{k+1} D_{ii} + \epsilon_N^{k+1}(N - \sum_{i=1}^N D_{ii}) + \gamma \sum_{i=N+1}^n D_{ii} \\
&= \sum_{i=1}^N \epsilon_i^{k+1} + \sum_{i=1}^N (\epsilon_N^{k+1} - \epsilon_i^{k+1})(1 - D_{ii}) + \gamma \sum_{i=N+1}^n D_{ii}.
\end{aligned}$$

As for any $1 \leq i \leq N$, $0 \leq D_{ii} \leq 1$ and $\epsilon_N^{k+1} \geq \epsilon_i^{k+1}$, we finally obtain

$$\text{Tr }(\widetilde{F}_k D) \geq \text{Tr }(\widetilde{F}_k D_{k+1}) + \frac{\gamma}{2}\|D - D_{k+1}\|^2.$$

The two statements of Lemma 2 follow. $\diamondsuit$

4 The Roothaan algorithm: why and how it fails

The Roothaan algorithm (also called *simple SCF* or *pure SCF* or *conventional SCF* in the literature) is the simplest fixed point procedure associated with the nonlinear eigenvalue problem (6). It consists in generating a sequence (D_k^{Rth}) in $\mathcal{P}$ satisfying

$$\left\{ \begin{array}{l} F(D_k^{Rth})C_{k+1} = C_{k+1}E_{k+1} \\ C_{k+1}^* C_{k+1} = I_N \\ D_{k+1}^{Rth} = C_{k+1}C_{k+1}^* \end{array} \right.$$

where $E_{k+1} = \text{Diag}(\epsilon_1^{k+1}, \cdots, \epsilon_N^{k+1})$, $\epsilon_1^{k+1} \leq \epsilon_2^{k+1} \leq \cdots \leq \epsilon_N^{k+1}$ being the N smallest eigenvalues of the linear eigenvalue problem

$$F(D_k^{Rth}) \cdot \phi = \epsilon\, \phi$$

and where the $n \times N$ matrix C_{k+1} collects N orthonormal eigenvectors of $F(D_k^{Rth})$ associated with ϵ_1^{k+1}, ϵ_2^{k+1}, ..., ϵ_N^{k+1}. The iteration procedure of the Roothaan algorithm can therefore be summarized by the diagram

$$D_k^{Rth} \quad \longrightarrow \quad \widetilde{F}_k = F(D_k^{Rth}) \quad \overset{\text{aufbau}}{\longrightarrow} \quad D_{k+1}^{Rth}.$$

The convergence properties of the Roothaan algorithm are not satisfactory: although the Roothaan algorithm sometimes numerically converges towards a solution to the HF equations, it frequently numerically oscillates between two states, none of them being solution to the HF equations. Numerical oscillation between two states means here that

$$D_{k+2}^{Rth} - D_k^{Rth} \longrightarrow 0, \qquad \text{but} \qquad D_{k+1}^{Rth} - D_k^{Rth} \not\longrightarrow 0.$$

The behavior of the Roothaan algorithm can be explained by introducing the auxiliary function

$$E(D, D') = \text{Tr }(hD) + \text{Tr }(hD') + \text{Tr }(G(D)\, D'),$$

which is symmetric since $\text{Tr }(G(D)\, D') = \text{Tr }(G(D')\, D)$, and which satisfies $E(D, D) = 2\, E^{HF}(D)$. Let us indeed minimize E alternatively with respect to each of the two arguments D and D':

$$D_1 = \text{arg inf}\left\{E(D_0, D), \quad D \in \mathcal{P}\right\},$$

$$D_2 = \text{arg inf}\left\{E(D, D_1), \quad D \in \mathcal{P}\right\},$$

$$D_3 = \text{arg inf}\left\{E(D_2, D), \quad D \in \mathcal{P}\right\},$$

$$\cdots$$

This minimization procedure is usually called *relaxation* in the mathematical literature. For the first two steps, we obtain

$$\begin{array}{rcl} D_1 & = & \text{arg inf}\left\{E(D_0, D), \quad D \in \mathcal{P}\right\} \\ & = & \text{arg inf}\left\{\text{Tr }(hD_0) + \text{Tr }(hD) + \text{Tr }(G(D_0)D), \quad D \in \mathcal{P}\right\} \\ & = & \text{arg inf}\left\{\text{Tr }(F(D_0)D), \quad D \in \mathcal{P}\right\} \\ & = & D_1^{Rth}, \end{array}$$

and, since E is symmetric on $\mathcal{P} \times \mathcal{P}$,

$$\begin{aligned} D_2 &= \arg\inf\Big\{E(D, D_1^{Rth}), \quad D \in \mathcal{P}\Big\} \\ &= \arg\inf\Big\{E(D_1^{Rth}, D), \quad D \in \mathcal{P}\Big\} \\ &= \arg\inf\Big\{\mathrm{Tr}\ (hD) + \mathrm{Tr}\ (hD_1^{Rth}) + \mathrm{Tr}\ (G(D_1^{Rth})D), \quad D \in \mathcal{P}\Big\} \\ &= \arg\inf\Big\{\mathrm{Tr}\ (F(D_1^{Rth})D), \quad D \in \mathcal{P}\Big\} \\ &= D_2^{Rth}. \end{aligned}$$

It follows by induction that the sequences generated by the relaxation algorithm on the one hand, and by the Roothaan algorithm on the other hand, are the same. The functional E, which decreases at each iteration of the relaxation procedure can therefore be interpreted as a Lyapunov functional of the Roothaan algorithm. This basic remark is the foundation of the proof of the following result.

Theorem 1. *Let $D_0 \in \mathcal{P}$ such that the Roothaan algorithm with initial guess D_0 is UWP. Then the sequence (D_k^{Rth}) generated by the Roothaan algorithm satisfies one of the following two properties*

- *either (D_k^{Rth}) numerically converges towards an aufbau solution to the HF equations*
- *or (D_k^{Rth}) numerically oscillates between two states, none of them being an aufbau solution to the HF equations.*

Proof. For any $k \in \mathbb{N}$, we deduce from Lemma 2 that

$$\mathrm{Tr}\ (F(D_{k+1}^{Rth})D_{k+2}^{Rth}) + \frac{\gamma}{2}\|D_{k+2}^{Rth} - D_k^{Rth}\|^2 \leq \mathrm{Tr}\ (F(D_{k+1}^{Rth})D_k^{Rth}).$$

Adding $\mathrm{Tr}\ (hD_{k+1}^{Rth})$ to both terms of the above inequality, we obtain

$$E(D_{k+1}^{Rth}, D_{k+2}^{Rth}) + \frac{\gamma}{2}\|D_{k+2}^{Rth} - D_k^{Rth}\|^2 \leq E(D_k^{Rth}, D_{k+1}^{Rth}).$$

We then sum up the above inequalities for $k \in \mathbb{N}$ and we get $\sum_{k\in\mathbb{N}} \|D_{k+2}^{Rth} - D_k^{Rth}\|^2 < +\infty$, which involves in particular that

$$D_{k+2}^{Rth} - D_k^{Rth} \longrightarrow 0.$$

Now, either $D_{k+1}^{Rth} - D_k^{Rth}$ converges to *zero* or it does not. In the former case, we deduce from the characterization of D_{k+1}^{Rth} by

$$\mathrm{Tr}\ (F(D_k^{Rth})D_{k+1}^{Rth}) = \inf\Big\{\mathrm{Tr}\ (F(D_k^{Rth})D), \quad D \in \mathcal{P}\Big\}$$

that

$$\mathrm{Tr}\ (F(D_k^{Rth})D_k^{Rth}) - \inf\Big\{\mathrm{Tr}\ (F(D_k^{Rth})D), \quad D \in \mathcal{P}\Big\} \longrightarrow 0.$$

Convergence towards an *aufbau* solution to the HF equations is thus established. In the latter case

$$\begin{aligned} \mathrm{Tr}\ (F(D_{2k}^{Rth})D_{2k}^{Rth}) - \inf\Big\{\mathrm{Tr}\ (F(D_{2k}^{Rth})D), \quad D \in \mathcal{P}\Big\} &= \mathrm{Tr}\ (F(D_{2k}^{Rth})D_{2k}^{Rth}) - \mathrm{Tr}\ (F(D_{2k}^{Rth})D_{2k+1}^{Rth}) \\ &\geq \frac{\gamma}{2}\|D_{2k}^{Rth} - D_{2k+1}^{Rth}\|^2 \not\longrightarrow 0. \end{aligned}$$

Convergence of (D_{2k}) towards an *aufbau* solution to the HF equations is therefore excluded; the same argument holds for (D_{2k+1}). $\diamondsuit$

Mimicking the proof of Theorem 2 (see section 5), it is easy to establish in addition that (D_{2k}, D_{2k+1}) converges up to an extraction to a critical point $(D, D') \in \mathcal{P} \times \mathcal{P}$ of the functional E which satisfies

$$\begin{cases} F(D')C = CE \\ C^*C = I_N \\ D = CC^* \\ F(D)C' = C'E' \\ C'^*C' = I_N \\ D' = C'C'^* \end{cases}$$

where E and E' are diagonal matrices collecting the smallest N eigenvalues of $F(D')$ amd $F(D)$ respectively. Besides, as $E(D_{2k}, D_{2k+1})$ is decreasing, the whole sequence (D_{2k}, D_{2k+1}) converges to (D, D') if this critical point is a strict (local) minimum. In this case, the alternatives are

- either (D, D') is on the diagonal of $\mathcal{P} \times \mathcal{P}$ (i.e. $D = D'$) and (D_k^{Rth}) converges towards an *aufbau* solution to the HF equations;
- or (D, D') is not on the diagonal of $\mathcal{P} \times \mathcal{P}$ (i.e. $D \neq D'$) and (D_k^{Rth}) oscillates between two states which are not *aufbau* solutions to the HF equations.

Both situations are represented on Figure 1.

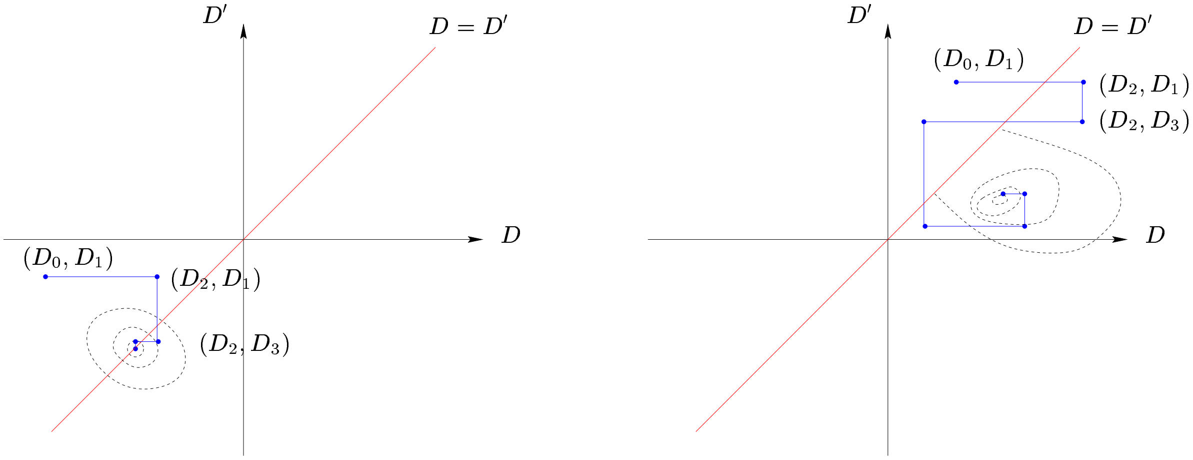

Figure 1: Minimization of E by relaxation: convergence towards a strict local minimum located on (resp. off) the "diagonal" leads to the convergence (resp. oscillations) of the Roothaan algorithm.

Oscillations can be observed even for simple chemical systems. As a matter of example, we have tested the Roothaan algorithm in the UHF setting for the atoms of the peridic table and for two sets of atomic orbitals, namely the gaussian basis sets 3-21G and 6-311++G(3df,3pd) (see [10]). The initial guess is obtained by diagonalization of the core hamiltonian. Calculations have been performed with Gaussian 98 [9]. The results are reported in Figure 2; they indicate that

1. Both alternatives (convergence *vs* oscillation) are met in practice.

2. Convergence towards a critical point of the HF problem which is not a global minimum can sometimes be observed.

3. For the same system, we can get convergence for one basis set and oscillation for another basis set.

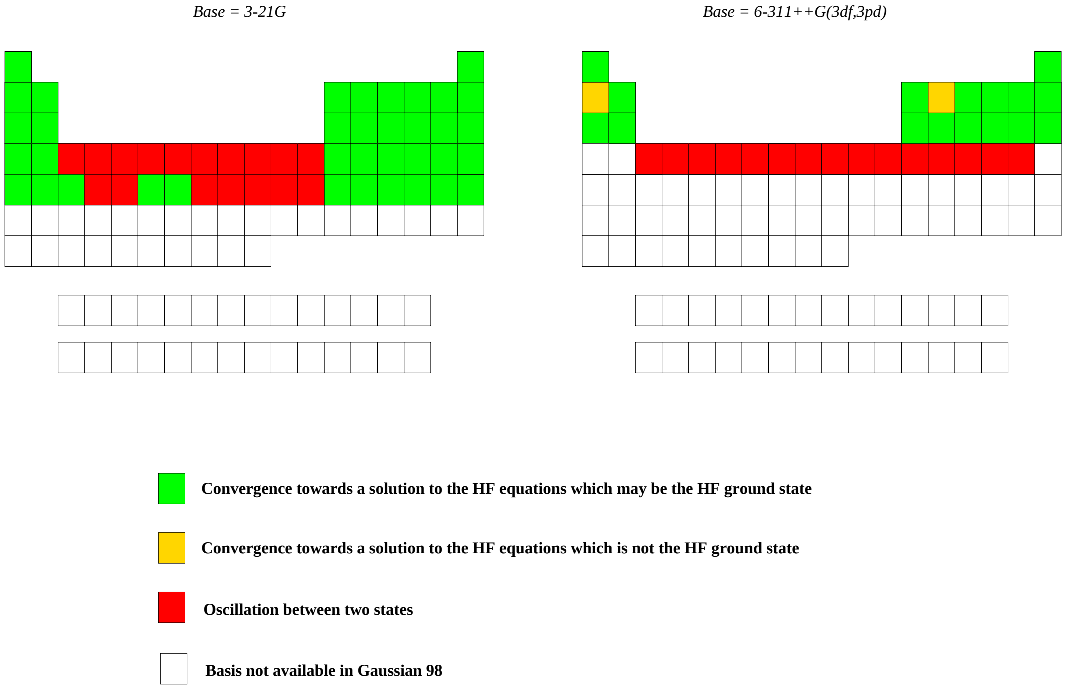

Figure 2: Searching the ground state of atoms with the Roothaan algorithm. Results are shown on the periodic table of the elements.

5 Level-shifting

The analysis developed in the previous section suggests to add to the functional E a penalization term E_p of the off-diagonal pairs (D, D') with $D \neq D'$ in order to enforce the critical points of the functional $E + E_p$ to lie on the diagonal of $\mathcal{P} \times \mathcal{P}$, which should ensure convergence towards a critical point of (5).

A simple penalization fonctional reads $E_p = b\|D - D'\|^2$, where b is a positive constant and where $\|\cdot\|$ denotes as above the Hilbert-Schmidt norm. Let us therefore set

$$E^b(D, D') = \text{Tr}\,(hD) + \text{Tr}\,(hD') + \text{Tr}\,(G(D)\,D') + b\,\|D - D'\|^2.$$

The relaxation algorithm associated with the minimization problem

$$\inf\left\{E^b(D, D'), \quad (D, D') \in \mathcal{P} \times \mathcal{P}\right\}$$

generates the sequence (D_k^b) defined by

$$D_k^b \quad \longrightarrow \quad \widetilde{F}_k = F(D_k^b) - b\,D_k^b \quad \overset{\text{aufbau}}{\longrightarrow} \quad D_{k+1}^b.$$

The sequence (D_k^b) can be identified with the sequence generated by the so-called level-shifting algorithm [23] with level-shift parameter b. The convergence of the level-shifting algorithm towards a (non necessarily *aufbau*) solution to the HF equations is mathematically guaranteed:

Theorem 2. *There exists a positive constant b_0 such that for any $D_0 \in \mathcal{P}$ and for any level-shift parameter $b \geq b_0$,*

1. *The sequence of the energies $E^{HF}(D_n^b)$ decreases towards some stationary value $\mathcal{E}$ of E^{HF}.*
2. *The sequence (D_n^b) numerically converges towards a solution to the HF equations.*

Proof. Let b_0 be a positive constant such that

$$\forall (D,D') \in \mathcal{M}(n,n) \times \mathcal{M}(n,n), \quad \text{Tr } \big(G(D-D')\cdot(D-D')\big) \leq b_0 \|D-D'\|^2. \quad (8)$$

Such a b_0 exists since $(d,d') \mapsto \text{Tr } (G(d)d')$ is a bilinear form on $\mathcal{M}(n,n)$. As E^b is symmetric on $\mathcal{P} \times \mathcal{P}$, we have for any $k \in \mathbb{N}$,

$$\begin{aligned} E^b(D_k^b, D_{k+1}^b) &= \inf \Big\{ E^b(D_k^b, D), \quad D \in \mathcal{P} \Big\} \\ &\leq E^b(D_k^b, D_k^b). \end{aligned}$$

A simple calculation shows that this inequality can be rewritten as

$$E^{HF}(D_{k+1}^b) - \frac{1}{2}\text{Tr } (G(D_{k+1}^b - D_k^b)\cdot(D_{k+1}^b - D_k^b)) + b\|D_{k+1}^b - D_k^b\|^2 \leq E^{HF}(D_k^b).$$

Therefore, for any $b \geq b_0$,

$$E^{HF}(D_{k+1}^b) + \frac{b}{2}||D_{k+1}^b - D_k^b||^2 \leq E^{HF}(D_k^b). \quad (9)$$

It follows that $(E^{HF}(D_k^b))$ is a decreasing sequence and that

$$\sum_{k=0}^{+\infty} ||D_{k+1}^b - D_k^b||^2 < +\infty.$$

The latter statement, which has been obtained by summing the inequalities (9) for $k \geq 0$, implies in particular that

$$D_{k+1}^b - D_k^b \longrightarrow 0.$$

As for any $k \in \mathbb{N}$,

$$[F(D_k^b) - bD_k^b, D_{k+1}^b] = 0,$$

it follows that

$$[F(D_k^b), D_k^b] = [F(D_k^b) - bD_k^b, D_{k+1}^b - D_k^b] \underset{k\to+\infty}{\longrightarrow} 0.$$

This concludes the proof of statement 2. As $\mathcal{P}$ is compact, we can extract from $(D_k^b)_{k\in\mathbb{N}}$ a subsequence $(D_{k_l}^b)_{l\in\mathbb{N}}$ which converges towards some $D \in \mathcal{P}$, such that $E^{HF}(D_{k_l}^b) \downarrow E^{HF}(D)$ and $[F(D), D] = 0$; $\mathcal{E} = \lim E^{HF}(D_k) = E^{HF}(D)$ is therefore a stationary value of the HF energy. $\diamondsuit$

The level-shift parameter b_0 implicitely defined by (8) is far from being optimal. Explicit and more refined estimates of shift parameters that garantee convergence are given in [3]. From a numerical viewpoint, it is important to choose not too large a shift parameter; otherwise the speed of convergence is very slow and the risk of converging towards a critical point whose energy is above that of the HF ground state is enhanced. This point, on which we will come back in the course of section 6, is illustrated by the numerical example reported on Figure 3 in which the level-shifting algorithm with various shift parameters has been used to compute the (doublet) UHF ground state of the Bromine atom in the gaussian basis 6-311++G(3df,3pd) (see [10]). In each case, the initial guess is computed by diagonalizing the core hamiltonian. Calculations have been performed with Gaussian 98 [9]. The algorithm oscillates for small shift parameters. For larger shift parameters, damped oscillations leading to convergence are observed. For very large shift parameter, the energy decreases at each iteration. Too large shift parameters have however to be excluded because they slow down the convergence (for $b = 30.0$ Ha, convergence towards the ground state is obtained after more than 200 iterations).

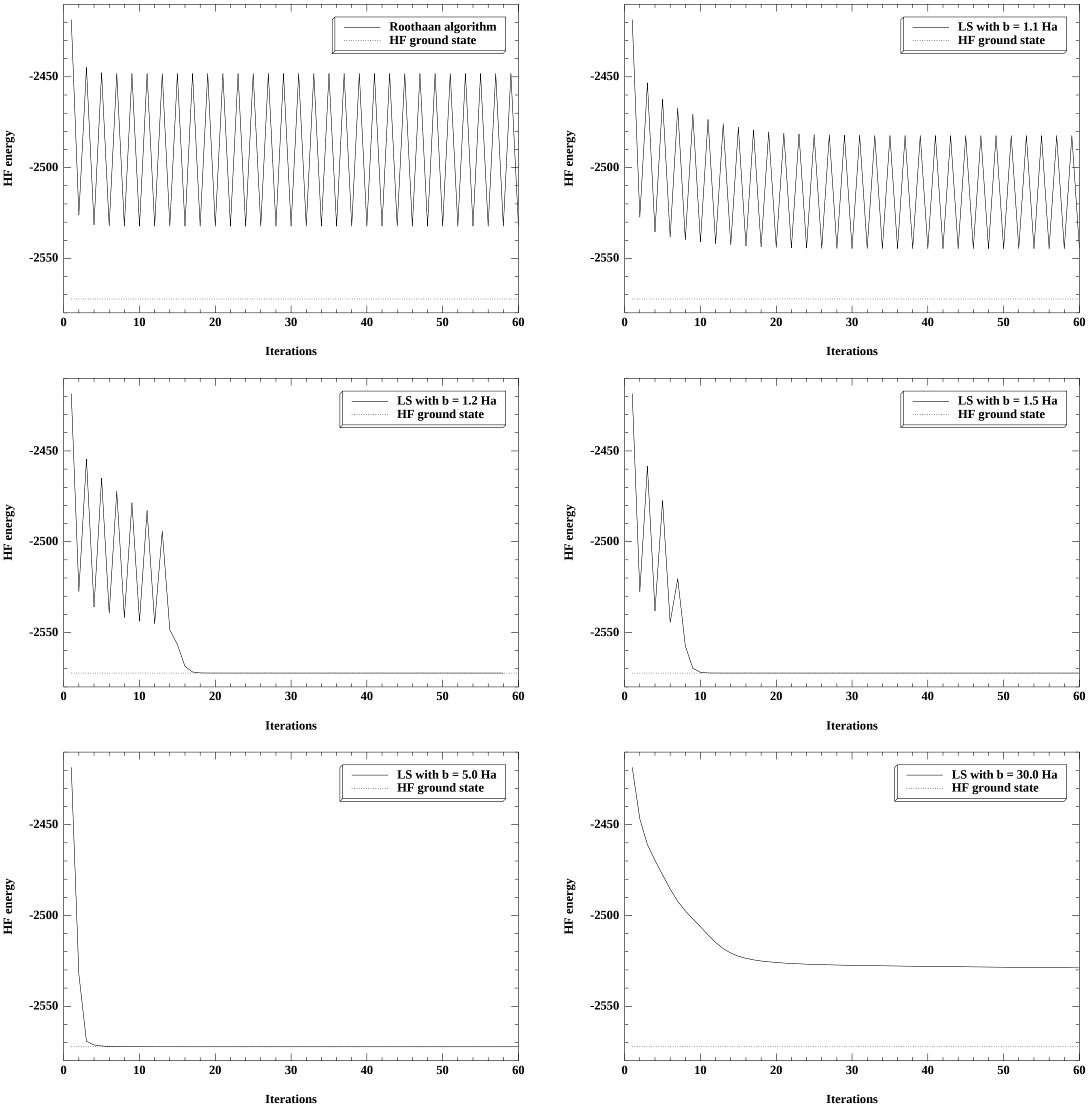

Figure 3: Calculation of the UHF ground state of the Bromine atom.

6 The Optimal Damping Algorithm

The present section is devoted to the mathematical study of the Optimal Damping Algorithm (ODA) which is the simplest representative of the class of Relaxed Constraints Algorithms (RCA) introduced in [4].

The ODA is defined by the following two-step iteration procedure

1. Diagonalize the current Fock matrix $\widetilde{F}_k = F(\widetilde{D}_k)$ and assemble the matrix $D_{k+1} \in \mathcal{P}$ by the *aufbau* principle;
2. Set $\widetilde{D}_{k+1} = \arg\inf \left\{ E(\widetilde{D}), \quad \widetilde{D} \in \text{Seg}[\widetilde{D}_k, D_{k+1}] \right\}$ where
$$\text{Seg}[\widetilde{D}_k, D_{k+1}] = \left\{ (1-\lambda)\widetilde{D}_k + \lambda D_{k+1}, \quad \lambda \in [0,1] \right\}$$
denotes the line segment linking $\widetilde{D}_k$ and D_{k+1}.

The procedure is initialized with $\widetilde{D}_0 = D_0$, $D_0 \in \mathcal{P}$ being a given initial guess.

The ODA thus generates two sequences of matrices:

- The principal sequence of density matrices $(D_k)_{k \in \mathbb{N}}$ which will be proved to numerically converge towards an *aufbau* solution to the HF equations;
- A secondary sequence $(\widetilde{D}_k)_{k \geq 1}$ of pseudo-density matrices which belong to the set
$$\widetilde{\mathcal{P}} = \left\{ \widetilde{D} \in \mathcal{M}(n,n), \quad \widetilde{D}^* = \widetilde{D}, \quad \text{Tr}\,(\widetilde{D}) = N, \quad \widetilde{D}^2 \leq \widetilde{D} \right\}$$
obtained from $\mathcal{P}$ by relaxing the nonlinear constraints $D^2 = D$.

The latter statement is a direct consequence of

Lemma 3. *The set $\mathcal{P}$ is convex,*

whose proof is postponed until the end of the present section. Indeed, $\widetilde{D}_0 = D_0 \in \mathcal{P} \subset \widetilde{\mathcal{P}}$ and, by induction, if $\widetilde{D}_k \in \widetilde{\mathcal{P}}$ then by convexity $\widetilde{D}_{k+1} \in \text{Seg}[\widetilde{D}_k, D_{k+1}] \subset \widetilde{\mathcal{P}}$ since $D_{k+1} \in \mathcal{P} \subset \widetilde{\mathcal{P}}$.

The properties of the ODA are put together in the following theorem.

Theorem 3. *For any initial guess D_0 for which the ODA is UWP,*

1. *The sequence $E(\widetilde{D}_k)$ decreases towards a stationary value of the HF energy.*
2. *The sequence $(D_k)_{k \in \mathbb{N}}$ converges towards an aufbau solution to the HF equations.*

As a first step towards the understanding of the ODA, let us consider $\widetilde{D}_k \in \widetilde{\mathcal{P}}$ and $D' \in \mathcal{P}$, and let us compute the variation of the HF energy on the line segment

$$\text{Seg}[\widetilde{D}_k, D'] = \left\{ (1-\lambda)\widetilde{D}_k + \lambda D', \quad \lambda \in [0,1] \right\}.$$

We obtain for any $\lambda \in [0,1]$,

$$E^{HF}((1-\lambda)\widetilde{D}_k + \lambda D') = E^{HF}(\widetilde{D}_k) + \lambda \text{Tr}\,(F(\widetilde{D}_k) \cdot (D' - \widetilde{D}_k)) + \frac{\lambda^2}{2} \text{Tr}\, \left(G(D' - \widetilde{D}_k) \cdot (D' - \widetilde{D}_k) \right).$$

The "steepest descent" direction, i.e. the density matrix D for which the slope $s_{\widetilde{D}_k \to D} = \text{Tr}\,(F(\widetilde{D}_k) \cdot (D - \widetilde{D}_k))$ is minimum, is given by the solution to the minimization problem

$$D = \arg\inf \left\{ \text{Tr}\,(F(\widetilde{D}_k) \cdot (D' - \widetilde{D}_k)), \quad D' \in \mathcal{P} \right\},$$

which also reads

$$D = \arg\inf \left\{ \mathrm{Tr}\,(F(\widetilde{D}_k)\cdot D'), \quad D' \in \mathcal{P} \right\}.$$

This is precisely the direction D_{k+1} obtained by the *aufbau* principle. The ODA can therefore be interpretated as a steepest descent algorithm in $\widetilde{\mathcal{P}}$. The practical implementation of the ODA is detailed in [4] for the RHF setting. The cost of one ODA iteration is approximatively the same as the cost of one iteration of the Roothaan algorithm (see [4] for details).

Figure 4 reports on a comparison between the ODA and the DIIS approaches for the calculation of the RHF ground state of the E form of n-methyl-2-nitrovinylamine (CH_3-NH-CH=CH-NO_2) in the basis 6-31G(d) (see [8]). The speed of convergence is estimated by computing the logarithm of the difference between the HF energy of the current density matrix and the (presumed) HF ground state energy. Calculations have been performed within Gaussian 98 [9]. The graph on the left hand side corresponds to an initial guess computed by a semiempirical method. In this case, both algorithms converge but the speed of convergence of the DIIS algorithm is higher. From a general viewpoint, numerical tests performed until now demonstrate that the ODA is efficient for performing the early iterations of the SCF procedure; when the sequence (D_k) has reached a neighbourhood of a critical point of the HF problem, convergence can be accelerated either by resorting to iterative subspace techniques or by switching to a quadratically convergent algorithm [4]. On the other hand, only the ODA converges for a more crude initial guess obtained by diagonalizing the core hamiltonian, as illustrated by the graph on the right hand side.

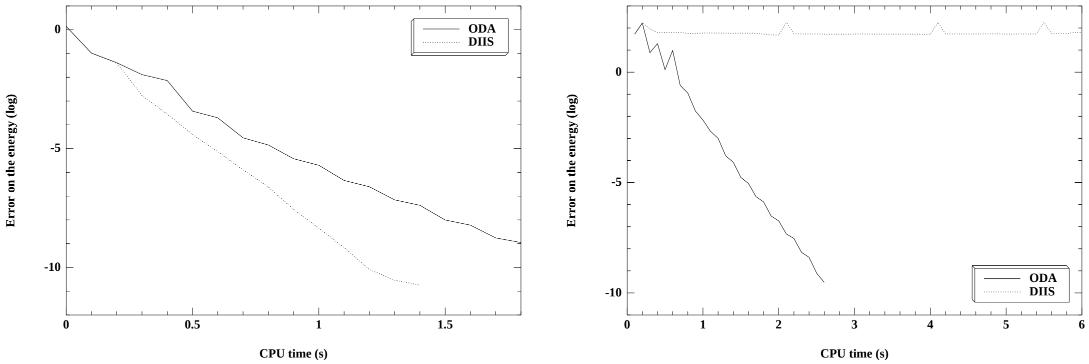

Figure 4: A comparison between the ODA and the DIIS algorithms: search for the RHF ground state of the E form of n-methyl-2-nitrovinylamine with an initial guess obtained by a semiempirical method (on the left hand side) and with the initial guess obtained by diagonalizing the core hamiltonian (on the right hand side).

Let us now detail the

Proof of Theorem 3. Let us denote by $\widetilde{F}_k = F(\widetilde{D}_k)$ and by $s_{k+1} = \mathrm{Tr}\,(\widetilde{F}_k(D_{k+1} - \widetilde{D}_k))$. In view of Lemma 2,

$$s_{k+1} = \mathrm{Tr}\,\left(\widetilde{F}_k D_{k+1}\right) - \mathrm{Tr}\,\left(\widetilde{F}_k \widetilde{D}_k\right) \leq -\frac{\gamma}{2}\|D_{k+1} - \widetilde{D}_k\|^2.$$

As above, let us denote by b_0 a positive constant such that

$$\forall (\widetilde{D}, \widetilde{D}') \in \mathcal{M}(n,n) \times \mathcal{M}(n,n), \qquad \mathrm{Tr}\,\left(G(\widetilde{D}' - \widetilde{D}) \cdot (\widetilde{D}' - \widetilde{D})\right) \leq b_0 \|\widetilde{D} - \widetilde{D}'\|^2.$$

For any $\lambda \in [0,1]$,

$$E^{HF}((1-\lambda)\widetilde{D}_k + \lambda D_{k+1}) \leq E^{HF}(\widetilde{D}_k) - \frac{\gamma}{2}\|D_{k+1} - \widetilde{D}_k\|^2\lambda + \frac{b_0}{2}\|D_{k+1} - \widetilde{D}_k\|^2\lambda^2.$$

Therefore

$$\begin{aligned} E^{HF}(\widetilde{D}_{k+1}) &= \inf\left\{E^{HF}((1-\lambda)\widetilde{D}_k + \lambda D_{k+1}), \qquad \lambda \in [0,1]\right\} \\ &\leq \inf\left\{E^{HF}(\widetilde{D}_k) - \frac{\gamma}{2}\|D_{k+1} - \widetilde{D}_k\|^2\lambda + \frac{b_0}{2}\|D_{k+1} - \widetilde{D}_k\|^2\lambda^2, \qquad \lambda \in [0,1]\right\} \\ &= E^{HF}(\widetilde{D}_k) - \alpha\|D_{k+1} - \widetilde{D}_k\|^2 \end{aligned}$$

with $\alpha = \gamma^2/8b_0$ if $\gamma \leq 2b_0$, $\alpha = (\gamma - b_0)/2$ otherwise. We then add up the above inequalities for $k \in \mathbb{N}$, and we get $\sum \|D_{k+1} - \widetilde{D}_k\|^2 < +\infty$, which implies that

$$D_{k+1} - \widetilde{D}_k \longrightarrow 0. \tag{10}$$

As $\widetilde{D}_{k+1} \in [\widetilde{D}_k, D_{k+1}]$, it follows that

$$\widetilde{D}_{k+1} - \widetilde{D}_k \longrightarrow 0,$$

and then that

$$D_{k+1} - D_k \longrightarrow 0.$$

Besides

$$\text{Tr }(F(\widetilde{D}_k)D_{k+1}) = \inf\left\{\text{Tr }(F(\widetilde{D}_k)D), \quad D \in \mathcal{P}\right\}. \tag{11}$$

Putting together (10) and (11), we finally obtain

$$\text{Tr }(F(D_{k+1})D_{k+1}) - \inf\left\{\text{Tr }(F(D_{k+1})D), \quad D \in \mathcal{P}\right\} \longrightarrow 0. \quad \diamondsuit$$

The following two points discuss the links between RCA and other algorithms like the level-shifting and the DIIS algorihms.

The first point concern the level-shifting algorithm. Let us use the ODA to minimize the penalized energy functional

$$E^b(D) = E^{HF}(D) - \frac{b}{2}\text{Tr }(D^2).$$

As for any $D \in \mathcal{P}$, $\text{Tr }(D^2) = \text{Tr }(D) = N$, the critical points of the minimization problem

$$\inf\left\{E^b(D), \quad D \in \mathcal{P}\right\} \tag{12}$$

are the same as those of the HF problem (5). On the other hand, for any $\widetilde{D} \in \widetilde{\mathcal{P}} \setminus \mathcal{P}$, $\text{Tr }(D^2) < N$: "interior" points are penalized. Let us denote by (D_k^b) and $(\widetilde{D}_k^b)$ the sequences generated by the ODA algorithm applied to (12):

1. Diagonalize the current Fock matrix $\widetilde{F}_k^b = F(\widetilde{D}_k^b) - b\widetilde{D}_k^b$ and assemble the matrix $D_{k+1}^b \in \mathcal{P}_N$ by the *aufbau* principle;

2. Set $\widetilde{D}_{k+1}^b = \text{arg inf}\left\{E(\widetilde{D}), \quad \widetilde{D} \in \text{Seg}[\widetilde{D}_k^b, D_{k+1}^b]\right\}$.

As for any $\lambda \in [0,1]$

$$\begin{aligned} E^b((1-\lambda)\widetilde{D}_k^b + \lambda D_{k+1}^b) &= E^b(\widetilde{D}_k^b) + \lambda \mathrm{Tr}\,(\widetilde{F}_k^b \cdot (D_{k+1} - \widetilde{D}_k)) \\ &\quad + \frac{\lambda^2}{2}\left(\mathrm{Tr}\,\left(G(D_{k+1}^b - \widetilde{D}_k^b)\cdot(D_{k+1}^b - \widetilde{D}_k^b)\right) - b\|D_{k+1}^b - \widetilde{D}_k^b\|^2\right), \end{aligned}$$

we obtain

$$s_{k+1} = \mathrm{Tr}\,(\widetilde{F}_k^b \cdot (D_{k+1} - \widetilde{D}_k)) \leq -\frac{\gamma}{2}\|D_{k+1}^b - \widetilde{D}_k^b\|^2$$

and for $b \geq b_0$

$$\mathrm{Tr}\,\left(G(D_{k+1}^b - \widetilde{D}_k^b)\cdot(D_{k+1}^b - \widetilde{D}_k^b)\right) - b\|D_{k+1}^b - \widetilde{D}_k^b\|^2 \leq 0.$$

For $b \geq b_0$, the function $\lambda \mapsto E^b((1-\lambda)\widetilde{D}_k^b + \lambda D_{k+1}^b)$ is therefore decreasing and concave on $[0,1]$; it follows that $\widetilde{D}_{k+1}^b = D_{k+1}^b$ which means that the ODA for minimizing (12) coincides with the level-shifing algorithm. This provides in particular another proof of the convergence of the level-shifting algorithm for large shift parameters b. Now, if D^b is an accumulation point of the sequence (D_k^b), we obtain by passing to the limit

$$s_\infty = \inf\left\{\mathrm{Tr}\,((F(D^b) - bD^b)\cdot(D - D^b)), \quad D \in \mathcal{P}\right\} = 0.$$

This implies that

$$D^b = \arg\inf\left\{\mathrm{Tr}\,((F(D^b) - bD^b)\cdot D), \quad D \in \mathcal{P}\right\},$$

and therefore that

$$[F(D^b) - bD^b, D^b] = [F(D^b), D^b] = 0,$$

but not necessarily that $D^b = \arg\inf\{\mathrm{Tr}\,(F(D^b)D), \quad D \in \mathcal{P}\}$: D^b is a solution to the HF equations that may not satisfy the *aufbau* principle. In addition, as $G(D) \geq 0$ for any $D \in \mathcal{P}$, we obtain

$$\begin{aligned} s_{k+1} &= \mathrm{Tr}\,(\widetilde{F}_k^b \cdot (D_{k+1}^b - D_k^b)) \\ &= \mathrm{Tr}\,(F(D_k^b)\cdot(D_{k+1}^b - D_k^b)) + \frac{b}{2}\|D_{k+1}^b - D_k^b\|^2 \\ &\geq -2\left|\inf\{\mathrm{Tr}\,(hD), \quad D \in \mathcal{P}\}\right| + \frac{b}{2}\|D_{k+1}^b - D_k^b\|^2. \end{aligned}$$

It results that

$$\|D_{k+1}^b - D_k^b\|^2 \leq \frac{2}{b}\left|\inf\{\mathrm{Tr}\,(hD), \quad D \in \mathcal{P}\}\right|.$$

The level-shifting algorithm can then also be interpreted as a trust region algorithm on the manifold $\mathcal{P}$ for which the radius of the trust region is bounded by $\delta = \frac{2}{b}\left|\inf\{\mathrm{Tr}\,(hD), \quad D \in \mathcal{P}\}\right|$. The larger the shift parameter b, the smaller the step $D_{k+1}^b - D_k^b$; this induces for large b a slow motion along a steepest descent path.

The second point is related to the DIIS algorithm. An attempt of improvement of the ODA consists, in the spirit of iterative subspace methods, in keeping in memory all (or some of) the density matrices computed at the previous steps and in minimizing the HF energy in the convex set generated by all the density matrices stored in memory:

1. Diagonalize $F(\widetilde{D}_k)$ and assemble the density matrix $D_{k+1} \in \mathcal{P}$ by the *aufbau* principle .

2. Set $\widetilde{D}_{k+1} = \arg\inf \left\{ E^{HF}(\widetilde{D}), \qquad \widetilde{D} = \sum_{i=0}^{k+1} c_i D_i, \quad 0 \leq c_i \leq 1, \quad \sum_{i=0}^{k+1} c_i = 1 \right\}$.

This algorithm is similar to Pulay's DIIS algorithm [21] except that in the DIIS algorithm, step 2 consists in minimizing the residual

$$\left\| \sum_{i=0}^{k+1} c_i [F(D_i), D_i] \right\|^2, \tag{13}$$

where $[.,.]$ denotes the commutator $[A,B] = AB - BA$, and where $\| \cdot \|$ denotes the Hilbert-Schmidt norm. Contrary to the RCA presented here, the DIIS algorithm may diverge: the residual (13) actually decreases at each step but it may vanish without the convergence is met.

Let us conclude this section with the

Proof of Lemma 3. Let $\widetilde{D}_1 \in \widetilde{\mathcal{P}}$ and $\widetilde{D}_2 \in \widetilde{\mathcal{P}}$. For any $0 \leq c_1, c_2 \leq 1$ such that $c_1 + c_2 = 1$,

$$\begin{aligned}
\widetilde{D}^2 &= (c_1\widetilde{D}_1 + c_2\widetilde{D}_2)^2 \\
&= c_1^2\widetilde{D}_1^2 + c_2^2\widetilde{D}_2^2 + c_1c_2(\widetilde{D}_1\widetilde{D}_2 + \widetilde{D}_2\widetilde{D}_1) \\
&= c_1\widetilde{D}_1 + c_2\widetilde{D}_2 + c_1(\widetilde{D}_1^2 - \widetilde{D}_1) + c_2(\widetilde{D}_2^2 - \widetilde{D}_2) \\
&\quad +(c_1^2 - c_1)\widetilde{D}_1^2 + (c_2^2 - c_2)\widetilde{D}_2^2 + c_1c_2(\widetilde{D}_1\widetilde{D}_2 + \widetilde{D}_2\widetilde{D}_1) \\
&= \widetilde{D} + c_1(\widetilde{D}_1^2 - \widetilde{D}_1) + c_2(\widetilde{D}_2^2 - \widetilde{D}_2) - c_1c_2(\widetilde{D}_1 - \widetilde{D}_2)^2
\end{aligned}$$

since $c_1^2 - c_1 = c_1(1 - c_1) = -c_1c_2 = c_2^2 - c_2$. Now, $\widetilde{D}_1^2 - \widetilde{D}_1 \leq 0$, $\widetilde{D}_2^2 - \widetilde{D}_2 \leq 0$ and $(\widetilde{D}_1 - \widetilde{D}_2)^2 \geq 0$. Consequently, $\widetilde{D}^2 \leq \widetilde{D}$. The other two constraints ($\widetilde{D} = \widetilde{D}^*$ and $\mathrm{Tr}\,(\widetilde{D}) = N$) being linear, it is clear that $\widetilde{D} \in \widetilde{\mathcal{P}}$. ♢

References

[1] V. Bach, E.H. Lieb, M. Loss and J.P. Solovej, *There are no unfilled shells in unrestricted Hartree-Fock theory*, Phys. Rev. Letters 72 (1994) 2981-2983.

[2] V. Bonačić-Koutecký and J. Koutecký, *General properties of the Hartree-Fock problem demonstrated on the frontier orbital model. II. Analysis of the customary iterative procedure*, Theoret. Chim. Acta 36 (1975) 163-180.

[3] E. Cancès and C. Le Bris, *On the convergence of SCF algorithms for the Hartree-Fock equations*, to appear in Math. Model. Num. Anal.

[4] E. Cancès and C. Le Bris, *Can we outperform the DIIS approach for electronic structure calculations*, Int. J. Quantum Chem., submitted.

[5] E. Cancès and C. Le Bris, *An efficient strategy to solve a nonlinear eigenvalue problem issued from electronic calculations in Quantum Chemistry*, J. Comput. Phys., submitted.

[6] J.-M. Combes, P. Duclos and R. Seiler, *The Born-Oppenheimer approximation*, in *Rigorous atomic and molecular physics*, G. Velo and A. Wightman (Eds), Plenum Press 1981.

[7] R. Fletcher, *Optimization of SCF LCAO wave functions*, Mol. Phys. 19 (1970) 55-63.

[8] J.B. Foresman and A. Frisch, *Exploring chemistry with electronic structure methods*, 2nd edition, Gaussian Inc., Pittsburgh PA 1996.

[9] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, B.B. Stefanov, G. liu, A. Liashenko, P. Piskorz, I. Kpmaromi, G. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B.G. Jpohnson, W. Chen, M.W. Wong, J.L. Andres, M. Head-Gordon, E.S. Replogle and J.A. Pople, Gaussian 98 (Revision A.7), Gaussian Inc., Pittsburgh PA 1998.

[10] A. Frisch and M.J. Frisch, *Gaussian 98 user's reference*, Gaussian Inc., Pittsburgh PA 1999.

[11] D.R. Hartree, *The calculation of atomic structures*, Wiley 1957.

[12] A. Igawa and H. Fukutome, *A new direct minimization algorithm for Hartree-Fock calculations*, Prog. Theor. Phys. 54 (1975) 1266-1281.

[13] J. Koutecký and V. Bonačić, *On convergence difficulties in the iterative Hartree-Fock procedure*, J. Chem. Phys. 55 (1971) 2408-2413.

[14] E.H. Lieb, *Bound on the maximum negative ionization of atoms and molecules*, Phys. Rev. A 29 (1984) 3018-3028.

[15] E.H. Lieb and B. Simon, *The Hartree-Fock theory for Coulomb systems*, Commun. Math. Phys. 53 (1977) 185-194.

[16] P.-L. Lions, *Solutions of Hartree-Fock equations for Coulomb systems*, Comm. Math. Phys. 109 (1987) 33-97.

[17] P.-L. Lions, *Hartree-Fock and related equations*, Nonlinear partial differential equations and their applications, Collège de France Seminar Vol. 9 (1988) 304-333.

[18] R. McWenny, *The density matrix in self-consistent field theory I. Iterative construction of the density matrix*, Proc. R. Soc. London Ser. A 235 (1956) 496-509.

[19] R. McWenny, *Methods of molecular Quantum Mechanics*, Academic Press 1992.

[20] A. Neumaier, *Molecular modeling of proteins and mathematical prediction of protein structure*, SIAM Rev. 39 (1997) 407-460.

[21] P. Pulay, *Improved SCF convergence acceleration*, J. Comp. Chem. 3 (1982) 556-560.

[22] C.C.J. Roothaan, *New developments in molecular orbital theory*, Rev. Mod. Phys. 23 (1951) 69-89.

[23] V.R. Saunders and I.H. Hillier, *A "level-shifting" method for converging closed shell Hartree-Fock wave functions*, Int. J. Quantum Chem. 7 (1973) 699-705.

[24] H.B. Schlegel and J.J.W. McDouall, *Do you have SCF stability and convergence problems?*, in *Computational Advances in Organic Chemistry*, Kluwer Academic, 1991, 167-185.

[25] T. Schlick, *Optimization methods in computational chemistry*, in *Reviews in Computational Chemistry, Vol. III*, K.B. Lipkowitz and D.B. Boyd (Eds.), VCH Publishers 1992.

[26] R. Seeger R. and J.A. Pople, *Self-consistent molecular orbital methods. XVI. Numerically stable direct energy minimization procedures for solution of Hartree-Fock equations*, J. Chem. Phys. 65 (1976) 265-271.

[27] R.E. Stanton, *The existence and cure of intrinsic divergence in closed shell SCF calculations*, J. Chem. Phys. 75 (1981) 3426-3432.

[28] R.E. Stanton, *Intrinsic convergence in closed-shell SCF calculations. A general criterion*, J. Chem. Phys. 75 (1981) 5416-5422.

[29] M.C. Zerner and M. Hehenberger, *A dynamical damping scheme for converging molecular SCF calculations*, Chem. Phys. Letters 62 (1979) 550-554.