CONVERGING SELF-CONSISTENT FIELD EQUATIONS IN QUANTUM CHEMISTRY – RECENT ACHIEVEMENTS AND REMAINING CHALLENGES *

KONSTANTIN N. KUDIN[1] AND GUSTAVO E. SCUSERIA[2]

Abstract. This paper reviews popular acceleration techniques to converge the non-linear self-consistent field equations appearing in quantum chemistry calculations with localized basis sets. The different methodologies, as well as their advantages and limitations are discussed within the same framework. Several illustrative examples of calculations are presented. This paper attempts to describe recent achievements and remaining challenges in this field.

Mathematics Subject Classification. 35P30, 65B99, 65K10, 81-08.

Received: October 10, 2005.

1. INTRODUCTION

In the course of several decades of dedicated effort, quantum chemistry techniques have become highly accurate and quite accessible to non-experts. Many users of modern quantum chemistry programs usually lack expertise in the numerical approaches which the methods are based on, and would very much prefer to use these computational programs in a fully "black box" manner, while still expecting qualitatively good predictions.

The Hartree-Fock (HF) method [15] or Density-Functional theory (DFT) [17,21] are among the least expensive yet still universally applicable and reliable approaches. These methods utilize orbitals or the density matrix as the basic variables. One of the most practical ways to solve the non-linear equations appearing in the HF or DFT method is to iterate some initial density (orbitals) until convergence, which means that a stationary point has been found and the input and output densities match [30]. This procedure is called the self-consistent field (SCF) iteration. Since in most cases the stationary point reached by the SCF iteration is actually a minimum, we will use these two terms interchangeably.

As an alternative to the SCF procedure, direct minimization techniques could be used. One of such, the original quadratically convergent self consistent field approach (QC-SCF) [1] is computationally rather expensive due to the need of evaluating the exact Hessian, at the very least in the form of a matrix-vector product. On

[1] Department of Chemistry and Princeton Institute for Science and Technology of Materials (PRISM), Princeton University, Princeton, NJ 08544, USA.

[2] Department of Chemistry, Rice University, Houston, Texas 77005-1892, USA. `guscus@rice.edu`

the other hand, recent work towards a direct minimization approach utilizing an approximate Hessian seems promising [37]. One should note here that the actual physical systems studied (atoms, molecules, and solids) are such that it is very much desired to find the global minimum in energy when it is non-degenerate, and a significant amount of effort can be expanded for a given solution in order to check that it is likely to be the global minimum [12, 33]. Of course, *fully* verifying that a given solution is the global minimum is impossible in practice. Even more importantly, for density functional theory (DFT) only the global minimum represents a physically relevant state of the system [17].

Here, we limit ourselves to discussing the approaches for accelerating the SCF convergence in the case of the simple iteration procedure. When using an iterative solver, a number of techniques can be employed in order to accelerate the SCF iteration process and enhance it convergence properties. The quantum chemistry community has tried many approaches helping to achieve convergence, with varying degree of success. Among the desired features of an acceleration approach is that it does not require any additional expensive intermediates not already available in the SCF procedure, and that the additional steps can be transparently incorporated into the algorithm. A number of techniques have attempted to tackle the problem from a purely mathematical point of view, for example by restricting the changes allowed in the density matrix from iteration to iteration. Then with the updates permitted small enough, the SCF iteration converges. In such methods, one has to manage the delicate balance between setting the adjustable parameters in such a way that the convergence is not slowed down too much, and yet can still be achieved.

Considering such a dilemma, parameter free acceleration techniques based on the solid mathematical understanding of the problem are significantly more attractive. For example, the direct inversion of the iterative subspace (DIIS) method introduced by Pulay [26, 27] uses the orbital rotation gradients in order to unambiguously compute the mixing coefficients. Due to extremely good practical behaviour, DIIS has perhaps become the most widely used method for accelerating the SCF convergence. Among the presently known techniques, DIIS is the most efficient approach for achieving convergence once the solution is close to a minimum [22]. Somewhat surprisingly, DIIS even works well if started quite far from a minimum in what we shall call "easy" systems. Why this is so is an interesting question that is reasonably well understood from a practical point of view while not fully explored from a rigorous mathematical perspective.

Another breakthrough came from the formal mathematical analysis of the convergence properties of the iterative Hartree-Fock equations carried out by Cancès [4, 6], who proposed the so called relaxed constraint algorithms (RCA) to address the issue of non-convergence. For a subset of the SCF methods, RCA has been proven to always find a minimum starting from *any* initial guess. The simplest implementation of RCA called the optimal damping algorithm (ODA), mixes two densities resulting in very robust behaviour demonstrated for a number of systems with convergence problems [7]. The logical evolution of the RCA type schemes is the energy-DIIS (EDIIS) [22], which is the most flexible among the RCA algorithms and allows mixing of several previous iterates. It is important to emphasize that while the mixing in the EDIIS does look mathematically simple, its true elegance stems from its link to the mathematical proof of convergence [7]. This particular point is especially highlighted by the fact that a damping scheme very similar in spirit to ODA had been described much earlier by Karlstrom [20]. However, lacking solid justification at the time of its introduction, this approach did not gain much popularity. Other versions of damping [29, 39] as well as level shifting schemes [32] were proposed in the same time frame, and without extensive empirical tests (not undertaken at the time) it was not possible to find out whether this particular version of damping had the very robust convergence properties that were later proven by Cancès and Le Bris [7].

At present, DIIS and EDIIS provide a powerful combination capable of achieving convergence quickly and reliably [22]. Nonetheless, this does not mean that there are no more practical problems in the field of SCF convergence in quantum chemistry. An important point that always needs to be kept in mind is that only the global minimum (or one of the lowest energy minima in the cases with degeneracies) are truly useful for the users of SCF methods. In many situations, the chemical systems are such that the global minimum is unique, and far lower in energy than the other ones, so it is sufficiently distinct and presents no challenges. On the other hand, the examples where the global minimum is hard to find are also quite abundant. Specifically,

the structures containing transition metals [9, 33] have many electronic levels clustered in a narrow physically important energy range, giving rise to many minima of similar energy. At present, such minima usually have to be thoroughly explored by the user, thus requiring a substantial amount of human effort [12]. Making such exploration more automatic would significantly improve the utility of the SCF methods for these systems.

Having discussed the current state of the SCF convergence acceleration in general terms, we now turn to a more detailed description. Specifically, in the upcoming sections, we will introduce the SCF problem in more precise terms, and discuss how various convergence acceleration schemes modify the iteration involved in order to accelerate or guarantee convergence.

2. Self consistent field equations

In this review we focus on the Hartree-Fock method, which is the simplest SCF approach. Later we comment on density functional approaches which also have the same structure.

In the Hartree-Fock method, one of the basic building blocks are one electron orbitals ψ_i, where for now i denotes the orbital index accounting both for the spatial and spin character of a given orbital. Then the Hartree-Fock problem can be formulated as

$$I^{HF} = \inf \left\{ E^{HF}(\psi_1, ..., \psi_{N_{tot}}), \psi_i \in H^1(\mathbf{R}^3), \int_{\mathbf{R}^3} \psi_i \psi_j^* = \delta_{ij} \right\}. \tag{1}$$

The Hartree-Fock energy is then

$$E^{HF}(\psi_1, ..., \psi_{N_{tot}}) = \frac{1}{2} \sum_{i=1}^{N_{tot}} \int_{\mathbf{R}^3} |\nabla \psi_i|^2 + \sum_{i=1}^{N_{tot}} \int_{\mathbf{R}^3} V |\psi_i|^2$$
$$+ \frac{1}{2} \int_{\mathbf{R}^3 \times \mathbf{R}^3} \frac{\rho(x)\rho(y)}{|x-y|} - \frac{1}{2} \int_{\mathbf{R}^3 \times \mathbf{R}^3} \frac{\tau(x,y)}{|x-y|} \tag{2}$$

where $\tau(x, y)$ and $\rho(x)$ are

$$\tau(x,y) = \sum_{i=1}^{N_{tot}} \psi_i(x)\psi_i^*(y), \rho(x) = \tau(x,x) \tag{3}$$

and V is the external (to electrons!) potential, such as the nuclear attraction.

To enforce the property of anti-symmetry, the ground state of the system is expressed as a single Slater determinant Ψ

$$\Psi(x_1, ..., x_{N_{tot}}) = \frac{1}{\sqrt{N_{tot}!}} \det(\psi_i(x_j)) \tag{4}$$

Most widely used quantum chemistry approaches expand ψ_i's in non-orthogonal atomic centered basis sets. Among various choices of the possible localized basis functions, Gaussians take a special place due to the availability of analytic formulae for the evaluation of the required integrals [3]. For a variety of reasons, it is convenient to separate the spatial and the spin indices of the orbitals, such as the orbitals are refered to as $\phi_i^\sigma \equiv \psi_{i'}$, with i now being the spatial index, and σ – the spin index. These spin orbitals are expressed as a sum of atom-centered non-orthogonal basis functions χ_μ

$$\phi_i^\sigma = \sum_{\mu=1}^{M} \chi_\mu C_{\mu i}^\sigma \tag{5}$$

where M is the total number of basis functions, and σ indicates the spin of the electron. Each spin orbital can be occupied by at most one electron. There are only two allowed values for the spin, α and β. In the *restricted* or *closed shell* Hartree-Fock (RHF) it is assumed that the spatial orbitals for both spins are the same, *i.e.* $\phi_i^\alpha = \phi_i^\beta$. Then, the numbers of electrons are $N_\alpha = N_\beta = N_{tot}/2$. In the following, for clarity of presentation, we focus on the RHF method only, and use the generic index σ to specify either of α or β spin states whenever it is useful.

Using the one electron orbital coefficients, one can define the density matrix as

$$D = C\mathcal{N}C^* \tag{6}$$

where $\mathcal{N}$ is a diagonal matrix with occupation numbers n_i on the diagonal, and we have dropped the spin index assuming either spin in the RHF case. In terms of occupations, the lowest N_α (and N_β) orbitals are assigned occupations 1, and the rest 0.

Both the Hartree-Fock energy and the density matrix are invariant with respect to orbital rotations within the occupied and virtual subspaces. In the density matrix representation, the restricted Hartree-Fock energy is expressed as

$$E^{RHF}(D) = 2(\text{Tr}\,(hD) + \frac{1}{2}\text{Tr}\,(G(D)D)) \tag{7}$$

where h is the core Hamiltonian consisting of the kinetic energy and nuclear attraction energy terms, and $G(D)$ is the 2 electron term consisting of the Coulomb and exchange contributions. The factor of 2 takes into account both electron spins α and β for a closed shell calculation. Note that in order to obtain the total energy, the nuclear repulsion energy needs to be added

$$E^{tot} = E^{RHF} + E^{nuc}. \tag{8}$$

The Fock matrix is then

$$F = \frac{1}{2}\frac{\delta E^{RHF}(D)}{\delta D} = h + G(D) \tag{9}$$

where factor $1/2$ takes into account the fact that we are dealing with the restricted case, for which formally $E^{RHF}(D) = E^{HF}(D_\alpha = D, D_\beta = D)$, and the Fock matrix is the variation with respect to only one of the density matrices while the second is kept fixed. One can either minimize the Hartree-Fock energy directly subject to the necessary constraints [1,37], or utilize an iterative procedure which in practice is usually more efficient. The SCF equations are then

$$FC = SC\epsilon \tag{10}$$
$$C^*SC = 1 \tag{11}$$

where F is the Fock matrix, S is the overlap of the non-orthogonal basis functions, C is the matrix of orbital coefficients, and ϵ are one electron orbital energies. Given sorted eigenvalues ϵ_i of the Fock matrix F_σ, the lowest N_σ orbitals are assumed to be occupied and have occupations of 1. This directly corresponds to the single determinant introduced in equation (4). The energy difference between $\epsilon(N_\sigma + 1)$ and $\epsilon(N_\sigma)$ is known as the HOMO-LUMO gap, *i.e.* the gap between the highest occupied (HOMO) and lowest unoccupied molecular orbitals (LUMO).

The number of electrons is directly related to the trace of the density matrix

$$\begin{aligned} \text{Tr}\,(D_\sigma S) &= N_\sigma,\ \text{or,} \\ \text{Tr}\,(\mathcal{D}_\sigma) &= N_\sigma \end{aligned} \tag{12}$$

where we have introduced the matrix $\mathcal{D}$ representing the density matrix in an orthonormal basis, related to the atomic basis D by

$$D = V\mathcal{D}V^*. \tag{13}$$

If $V = C$ for the current iteration (*i.e.* basis for the self-consistent field Eq. (10)), then $\mathcal{D}$ is diagonal as was in equation (6). An important property of the density matrix obtained from a single Slater determinant is known as idempotency

$$\begin{aligned} DSD &= D, \text{ or,} \\ \mathcal{D}\mathcal{D} &= \mathcal{D}. \end{aligned} \tag{14}$$

A given acceleration technique may or may not maintain idempotency in certain intermediate density matrices while approaching convergence. Despite some recent mathematically unsubstantiated claims to the contrary [36], lack of idempotency in the intermediates does not imply any deficiency in the method and may in fact be a desired goal as in the RCA case [7].

A necessary condition for a stationary point in the SCF equations is that the density matrix and its corresponding Fock matrix commute

$$\begin{aligned} FDS - SDF &= 0, \text{ or,} \\ \mathcal{F}\mathcal{D} - \mathcal{D}\mathcal{F} &= 0. \end{aligned} \tag{15}$$

The Fock matrix $\mathcal{F}$ in the orthonormal basis V can obtained as

$$\mathcal{F} = V^*FV. \tag{16}$$

The commutator $[\mathcal{F}, \mathcal{D}]$ represents the energy gradient with respect to the orbital rotations, and thus naturally vanishes at a stationary point. We will revisit this question more extensively in the context of the DIIS method.

Having set up the basic intermediates, we can now introduce the Roothan algorithm which is the simplest approach to the solution of the SCF equations

$$D_n \to F(D_n) \xrightarrow[aufbau]{diagonalization} D_{n+1}. \tag{17}$$

In the above equation one diagonalizes $F(D_n)$ and then forms D_{n+1} according to the *aufbau* principle: given the sorted eigenvalue ϵ_i spectrum of the current Fock matrix $F(D_n)$, one occupies the lowest in energy N_σ orbitals for the next iteration. These are also known as *aufbau* occupations, and the electronic states where this principle is violated (*i.e.* $\epsilon(N_\sigma + 1) < \epsilon(N_\sigma)$) are called states with *aufbau* violations. One can prove that for the *unrestricted* or *generalized* HF method a proper *aufbau* solution always exists [23].

This procedure for solving the SCF equations is also known as the *fixed point* iteration. The Roothan algorithm achieves convergence in well behaved cases. Its typical non-convergence behaviour are oscillations between two densities neither of which is an SCF solution [6]. Thus, acceleration techniques are not a luxury, but rather a necessity. Studies have linked poor SCF convergence with a small HOMO-LUMO gap and the oscillations described above to typical *attractors* appearing when solving non-linear equations by simple interation [25]. As a general note we should add here that limiting the changes in the density matrix from cycle to cycle by different means makes the Roothan method more robust at the possible expense of extra SCF iterations. The main issue then is to find an acceleration method that not only achieves convergence, but also minimizes the number of iterations required while reaching preferably a minimum of low energy.

3. ACCELERATION APPROACHES

In this section, we discuss the structure and the foundations of the various acceleration techniques usually employed to help achieve SCF convergence. For the benefit of further discussion below, we slightly rewrite the Roothan algorithm in a more extended way

$$D_n \rightarrow F(D_n) \rightarrow \left[\tilde{F}_n = F(D_n) \right] \xrightarrow[aufbau]{diagonalization} D_{n+1}. \tag{18}$$

In the above equation, we have introduced the matrix $\tilde{F}_n$, which now can differ from $F(D_n)$. Most acceleration techniques are concerned with forming the matrix $\tilde{F}_n$ by using the information about the Fock and density matrices from previous SCF iterations. Often (but not always) there exists a matrix $\tilde{D}_n$ such that $\tilde{F}_n = F(\tilde{D}_n)$.

A rather useful property of the Fock matrix is that

$$F\left(\sum_i c_i D_i \right) = \sum_i c_i F(D_i) \tag{19}$$

for any set of c_i's that satisfy $\sum c_i = 1$. In fact, as we discuss below, a number of acceleration techniques do exploit this property in order to avoid recomputing the Fock matrix for density matrix combinations. In the upcoming subsections, we will briefly describe several popular techniques, roughly in the order that they were introduced.

3.1. Damping

Simple damping mixes output density with the input one, thus limiting the density changes from cycle to cycle [15].

$$\tilde{D}_{n+1} = (1 - \lambda)\tilde{D}_n + \lambda D_{n+1} \tag{20}$$
$$\tilde{F}_{n+1} = F(\tilde{D}_{n+1}). \tag{21}$$

Written in this way, the input density matrix $\tilde{D}_{n+1}$ to the Fock construction procedure is not idempotent. It is straightforward to rewrite the damping in a way that generates an idempotent sequence of the density matrices.

$$D_n \rightarrow \left[\tilde{F}_n = (1 - \lambda)\tilde{F}_{n-1} + \lambda F(D_n) \right] \rightarrow D_{n+1}. \tag{22}$$

Such restructuring does not change the nature of the algorithm. The mixing coefficient λ can be either fixed or chosen based on some criteria.

By limiting the changes in $\tilde{F}_n$ between cycles, the procedure can be reasonably successful in helping to achieve convergence in some cases where the basic Roothan algorithm fails [18]. Overall, the mixing coefficient can be chosen based on a number of criteria, introducing certain arbitrariness. Moreover, in such a formulation, there is no clear physical meaning attached to λ. While relaxed constrained algorithms (RCA) do utilize the same concept, we will discuss them in a separate subsection due to the well defined convergence properties and lack of any arbitrariness.

3.2. Level shift approach

The level shift method introduced by Saunders and Hillier [32] aims to limit the rotations between the occupied and virtual subspaces by generating $\tilde{F}_n$ from $F(D_n)$ according to the expression

$$\tilde{F}_n = F(D_n) - \mu D_n. \tag{23}$$

This transformation is such that there is no $\tilde{D}_n$ that corresponds to this $\tilde{F}_n$. The level shift parameter μ makes the occupied space (D_n) more likely to be occupied again by artificially lowering the energy of the occupied orbitals by μ. Thus, by construction, the sequence of *idempotent* D_n's contains no large changes between consecutive iterations. This feature makes the level shift technique extremely prone to converge to a nearby local minimum since any exploration of the orbital space is mostly inhibited. By making the level shift parameter sufficiently large, one can force the changes between SCF cycles to be arbitrarily small, so that the SCF will always converge to a nearby local minimum. There are mathematical proofs available that this is indeed so [6]. Of course, safe level shift values may require a large number of SCF cycles to achieve convergence. So an important question with the level shift method is how to choose the parameter μ in a way that the convergence is optimal. In order to optimize the level shift behavior, Mitin has discussed the idea of dynamical level shift [24]. Recently, Thogersen *et al.* [35,36] as well as others [13] have revisited the method by introducing a trust radius on the allowed density matrix changes driving the choice of μ.

An interesting idea is to use the method in a reverse way, *i.e.* to increase the mixing between the occupied and virtual subspaces by a negative level shift value. Such technique was proposed before the advent of reliable acceleration approaches [2], and should perhaps be reinvestigated for its utility in combination with the more robust methods.

3.3. Direct inversion of the iterative subspace (DIIS)

As mentioned in the Introduction, DIIS [26,27] is the method of choice for SCF convergence when close to a minimum. In the math literature, the technique similar to DIIS is known as the generalized minimum residue (GMRES) method [31]. The Fock matrix $\tilde{F}_n$ for the diagonalization is obtained as $F(\sum c_i D_i)$ with underlying $\tilde{D}_n$ that – unlike level shift – does exist. In one of its widely used forms, the DIIS method evaluates the mixing coefficients for the previous iterates by minimizing the function

$$f^{DIIS}(c_0, ..., c_n) = \left\| \sum_{i=0}^{n} c_i [\mathcal{F}_i, \mathcal{D}_i] \right\|^2 \tag{24}$$

where $\| \cdot \|$ denotes the Frobenius norm and $[\mathcal{F}_i, \mathcal{D}_i]$ is the commutator of Fock and density matrices in some orthonormal basis. This orthonormal basis should of course be the same for all the commutators in the above expression. Original references [26,27] have briefly outlined why the DIIS scheme accelerated convergence. Here, we would like to offer a more elaborate explanation instead, such that otherwise hidden details of the DIIS *with commutator matrices* is revealed.

An idempotent parametrization of density matrix $\mathcal{D}_a$ in some orthonormal basis can be obtained as follows [16, 36]

$$\mathcal{D}_a(X) = \mathrm{e}^{-X} \mathcal{D}_a \mathrm{e}^X = \mathcal{D}_a + [\mathcal{D}_a, X] + \frac{1}{2}[[\mathcal{D}_a, X], X] + ... \tag{25}$$

where X is an anti-symmetric matrix, and e^X is its exponential.

As discussed earlier (Eq. (9)), the gradient of the RHF energy $E^{RHF}(D_i)$ is simply $2F(D_i) = 2F_i$ at any density D_i. On the other hand, the gradient of the RHF energy parametrized *via* X as $E_i^{RHF}(X) = E^{RHF}(\mathcal{D} = \mathrm{e}^{-X} \mathcal{D}_i \mathrm{e}^X)$ is $[\mathcal{F}_i, \mathcal{D}_i]$ at $X = 0$, where $\mathcal{F}_i = \mathcal{F}(\mathcal{D}_i)$. To simplify further notation, we write

$$g(\mathcal{D}_i) = \frac{\partial E_i^{RHF}(X)}{\partial X} \mid_{X=0} = [\mathcal{F}_i, \mathcal{D}_i]. \tag{26}$$

Now, let us assume that the sequence of iterative densities D_n resides in the basin of some self-consistent density $\hat{D}$ ($\hat{\mathcal{D}}$ in an orthonormal basis). The gradient of $\hat{\mathcal{D}}$ is zero (as per Eq. (26)) since it is a stationary point and thus the first order variation should be zero. At this stage we can formally represent each known $\mathcal{D}_i$ as

$$\mathcal{D}_i = \mathrm{e}^{-X_i} \hat{\mathcal{D}} \mathrm{e}^{X_i} \tag{27}$$

where both X_i and $\hat{\mathcal{D}}$ are unknowns. Using this relation, we can express $g(\mathcal{D}_i)$ as a Taylor expansion of the gradient near $\hat{\mathcal{D}}$

$$
\begin{aligned}
g(\mathcal{D}_i(X)) \mid_{X=0} &= g(\mathrm{e}^{-(X_i+X)}\hat{\mathcal{D}}\mathrm{e}^{(X_i+X)}) \mid_{X=0} \\
&= g(\hat{\mathcal{D}}) + H(\hat{\mathcal{D}})X_i + O(X_i^2) \\
&= H(\hat{\mathcal{D}})X_i + O(X_i^2)
\end{aligned}
\tag{28}
$$

where $H(\hat{\mathcal{D}})$ is the Hessian matrix with respect to X. As mentioned earlier, the gradient $g(\hat{\mathcal{D}})$ is zero.

Now, while neither $\hat{\mathcal{D}}$ nor X_i's are known, we still would like to find some combined density which would be the closest to the stationary $\hat{\mathcal{D}}$. In order to eliminate the unkowns, we introduce $\tilde{X}$

$$
\tilde{X} = \sum_i c_i X_i.
\tag{29}
$$

The corresponding density matrix is then

$$
\mathcal{D}(\tilde{X}) = \mathrm{e}^{-\tilde{X}}\hat{\mathcal{D}}\mathrm{e}^{\tilde{X}}.
\tag{30}
$$

Utilizing equation (25), expansion of $\mathcal{D}(\tilde{X})$ gives

$$
\begin{aligned}
\mathcal{D}(\tilde{X}) &= \hat{\mathcal{D}} + \sum_i c_i[\hat{\mathcal{D}}, X_i] + O(X_i X_j) \\
&= (1 - \sum c_i)\hat{\mathcal{D}} + \sum_i c_i \mathcal{D}_i + O(X_i X_j)_{i \neq j}.
\end{aligned}
\tag{31}
$$

Thus, by requiring $[\sum c_i = 1]$, we can eliminate to the lowest order the unknown stationary $\hat{\mathcal{D}}$ from the above expression. Then a reasonable approximation for $\mathcal{D}(\tilde{X})$ is $\sum_i c_i \mathcal{D}_i$. What remains to be found are the mixing coefficients c_i's. The electronic gradient for $\mathcal{D}(\tilde{X})$ is

$$
\begin{aligned}
g(\mathcal{D}(\tilde{X})) &= \sum_i c_i H(\hat{\mathcal{D}})X_i + O(X_i X_j) \\
&= \sum_i c_i g(\mathcal{D}_i) + O(X_i X_j)_{i \neq j} \\
&\sim c_i \sum [\mathcal{F}_i, \mathcal{D}_i]
\end{aligned}
\tag{32}
$$

where we have employed both equations (28) and (26). The gradient defined by equation (32) can be minimized subject to the constraint on the sum of c_i's, leading to equation (24). Within the approximations adopted, this should give $\tilde{D}_n$ closest to $\hat{D}$, for which $\tilde{X}$ is the smallest. An interesting feature here is the implicit use of the Hessian matrix *via* equation (32). With this, one could argue that the DIIS with commutators is in fact a second order (quasi-Newton) technique. A somewhat similar analysis of the DIIS was recently given in reference [36]. Since $\mathcal{D}(\tilde{X})$ was originally constructed to be idempotent, it is rather desirable that the *extrapolated* density matrix $\sum_i c_i \mathcal{D}_i$ is also close to being idempotent such that the asumptions hold.

From the principle underlying the DIIS method, it also becomes clear that in theory the method should only work well for a sequence of density matrices in the vicinity of the *same* minimum such that the expansions described by equations (31) and (32) are valid. In practice, DIIS turns out to be even more robust than one would expect. Otherwise, to aid the situation, approaches like level shift can be used to generate a sequence of density matrices in the basin of the same minimum. Thus it is not surprising that the combination of these methods has been successfully used for achieving SCF convergence in many difficult situations.

3.4. **Relaxed constraints algorithms (RCA)**

The main idea behind the RCA [7] is to explicitly make the intermediate density matrix non-idempotent yet still belonging to a convex set

$$\tilde{D}S\tilde{D} \leq \tilde{D} \tag{33}$$

This is identical to allowing occupation numbers n_i to vary continuously between 0 and 1. An earlier version of the RCA – optimal damping algorithm (ODA) [7] – utilized just two matrices $\tilde{D}_n$ and D_{n+1}. The non-idempotent density matrix $\tilde{D}_n$ accumulates the history of the previous iterations, while an idempotent matrix D_{n+1} derived from diagonalizing $\tilde{F}_n$ provides a new steepest descent search direction. Otherwise, the mixing expression is identical to the generic damping represented by equation (20), of course now with an optimized mixing coefficient λ at each iteration.

The choice of the coefficients comes from the minimization of the Hartree-Fock energy, which is an exact quadratic function of the mixing coefficients [7, 20]

$$\begin{aligned}
E^{RHF}\left((1-\lambda)\tilde{D}_n + \lambda D_{n+1}\right) = {} & E^{RHF}(\tilde{D}_n) \\
& + 2\lambda \mathrm{Tr}\left(F(\tilde{D}_n) \cdot (D_{n+1} - \tilde{D}_n)\right) \\
+ \lambda^2 \mathrm{Tr}\left((F(D_{n+1}) - F(\tilde{D}_n)) \cdot (D_{n+1} - \tilde{D}_n)\right). &
\end{aligned} \tag{34}$$

The inequality on the density matrix $\tilde{D}_n$ specified by equation (33) is maintained by requiring that $0 \leq \lambda \leq 1$. Since the Hartree-Fock energy functional has no local minimum on the convex set of non-idempotent density matrices [7], the ODA can be mathematically shown to always converge to an idempotent density matrix solution with no *aufbau* violations for the generalized Hartree-Fock and unrestricted Hartree-Fock methods [5]. For the *restricted* Hartree-Fock this is not true, and actual counterexamples have already been found[1].

As mentioned in the Introduction, an energy minimization for current D_{n+1} and previous D_n density matrices similar to equation (34) was proposed by Karlstrom years ago [20]. However, the superior convergence properties of such an approach did not become evident until the formal work of Cancès and Le Bris [7].

One can observe that in the ODA the non-idempotent matrix $\tilde{D}_n$ contains a mix of previous D_i's with coefficients rigidly set by previous iterations [22]. By extending the ODA to any number of density matrices, one can eliminate the non-idempotent matrix $\tilde{D}_n$ and work with the previous iterates directly. Denoting $f^{EDIIS} = E^{HF}\left(\sum_{i=0}^{k} c_i D_i\right)$, the Hartree-Fock energy can be written as

$$f^{EDIIS}(c_0, ..., c_k) = \sum_{i=0}^{k} c_i E^{HF}(D_i) - \frac{1}{2} \sum_{i,j=0}^{k} c_i c_j \mathrm{Tr}\left((F_i - F_j) \cdot (D_i - D_j)\right) \tag{35}$$

which is the working formula of the EDIIS (energy-DIIS) approach. The convexity assumption on the *interpolated* density $\tilde{D}_{n+1}$ is maintained by making the coefficients satisfy the relation $0 \leq c_i \leq 1$. Naturally, for just two matrices this expression can be shown to be identical to equation (34). The EDIIS permits full flexibility in how previous iterates are mixed, and thus is the fastest version of the relaxed constraints algorithms to date.

An important point that relates to EDIIS is that in early iterations the consecutive (and idempotent!) density matrices D_n are not forced to be similar in any way as compared to the level shift method. Thus EDIIS has a chance to explore a larger electronic subspace while eventually settling on a specific local minimum. This is indeed the behaviour observed in practice.

---

[1]K.N. Kudin, unpublished.

3.5. **Other approaches**

So far we have assumed that the computational cost for the SCF problem is dominated by the Fock matrix construction. This is true for the majority of systems currently studied. Therefore, acceleration techniques specifically aim at minimizing the number of Fock matrix constructions. However, as the size of systems continues to increase, linear scaling approaches can be employed for Fock matrix construction, but the diagonalization step becomes the dominant step of the SCF calculation [34]. Then one must switch to a linear scaling alternative to diagonalization. Quite interestingly, some of the alternatives like the conjugate gradient density matrix search (CGDMS) enhance the convergence properties of the fixed point SCF algorithm. Since in the CGDMS method the new density is obtained by updating the old one, the discontinuous changes in the occupation numbers between the successive iterations do not occur, which stabilizes the SCF convergence [9].

Various orbital occupation broadening techniques have been shown to have positive influence on the SCF convergence. In these methods, orbital occupations are broadened in early iterations utilizing a fictitious temperature parameter and a suitable broadening function such as Fermi-Dirac distribution. Here again, density matrix changes from cycle to cycle are more restricted than otherwise, so that the SCF procedure may converge better [28].

4. Beyond Hartree-Fock: density functional theory

The Hartree-Fock method only accounts for the average interaction between electrons, and completely misses electronic correlation – the fact that in physical systems the electrons "avoid" each other by "moving" in a correlated way. Thus the true electronic repulsion (and the energy of the system) is in fact lower than what is computed by the Hartree-Fock method. Density Functional Theory (DFT) provides (at least in principle) an exact way to include these effects [17].

The Kohn-Sham approach [21] to Density Functional Theory (DFT) is a practical tool to compute the exchange and correlation energies in a way comparable in cost and simplicity to the Hartree-Fock method, yet resulting in a substantially more accurate approach [19]. Moreover, even though the Kohn-Sham equations are based on a quite different physical principle, they give rise to SCF equations mathematically very similar to those of the Hartree-Fock method. In the density matrix representation the electronic energy for a closed shell system is

$$E^{KS}(D) = 2(\text{Tr}\,(hD) + \frac{1}{2}\,\text{Tr}\,(J(D)D)) + E_{xc}(D) \tag{36}$$

where $J(D)$ is the electron-electron Coulomb repulsion term, and E_{xc} accounts for the additional exchange-correlation potential. The (restricted) Kohn-Sham matrix is

$$F = \frac{1}{2}\frac{\delta E^{KS}(D)}{\delta D} = h + J(D) + \frac{1}{2}\frac{\delta E_{xc}(D)}{\delta D} \tag{37}$$

where again both factors of 1/2 account for the spin-restricted version of the method because $E^{RKS}(D) = E^{KS}(D_\alpha = D, D_\beta = D)$, and $E_{xc}(D) = E_{xc}(D_\alpha = D, D_\beta = D)$, and the Fock matrix is the variation with respect to only one of the densities. While being extremely important physically, both E_{xc} and $\delta E_{xc}(D)/\delta D$ are smaller numerically than the other terms contributing to the energy and Kohn-Sham matrices. Thus convergence issues arising in the DFT calculations are in practice mostly similar to those in Hartree-Fock.

The same acceleration approaches (including EDIIS)[2] can be used rather successfully for a variety of systems. While formally a Kohn-Sham matrix for a sum of density matrices is no longer a sum of respective Kohn-Sham matrices (as in the HF method, Eq. (19)), in practice, such an approximation is quite acceptable. So, just like in Hartree-Fock, acceleration schemes usually do not recompute the Kohn-Sham matrix, but simply use a corresponding linear combination of the previous Kohn-Sham matrices.

An important issue arising in DFT calculations is that the true exchange-correlation functional is not known, so various approximations have to be employed in practice. For a subset of the functionals commonly used, there are systems where integer occupied solutions without *aufbau* violations simply do not exist. *Extended* Kohn-Sham model allow the occupations to be fractional [11] (the model with integer occupations is called *standard* Kohn-Sham). In calculations with the extended Kohn-Sham model, there is a number of low energy levels with occupation 1, then there is a group of levels with the same energy and fractional occupations between 0 and 1, after which empty levels with higher energy follow [5].

In such systems it is important to recognize that a proper *aufbau* solution is not achievable, and handle the situation appropriately. The RCA methods are well suited for detecting such cases [5]. One possible approach to obtain a solution is to use orbital mapping [38] or level shift to force an integer occupied solution with *aufbau* violations. Another possibility is to obtain the approximate fractional occupation numbers by utilizing the RCA [5], and then fully optimize them *via* the quadratically convergent (QC-SCF) approach for fractional occupations [8]. The RCA can occasionally lead to a solution with fractional occupations even for the *restricted* Hartree-Fock method (footnote[1]). However, such a situation is readily recognizable.

5. EXAMPLES

Let us take a look now at how the various acceleration techniques work in practice. This is a very useful exercise in order to get a feel for how many SCF cycles are usually required, and determine what kind of convergence properties any new technique should possess in order to beneficially complement or even replace any of the existing approaches.

Our tests systems will include iso-propyl-alcohol [$(CH_3)_2CHOH$], benzene (C_6H_6), and NiC. The first two molecules seem to have a single global minimum well separated from any other local minima. Their geometries were here optimized at the RHF/6-31G(d) level of theory. On the other hand, the NiC molecule at the internuclear distance of 1.8 Å has multiple electronic minima of similar energy. In all cases, the SCF calculations were carried out at the RHF/6-31G(d) level of theory using the *Gaussian03* program (*G03*) [14]. The basis set includes the d functions in Cartesian form, thus utilizing 6 of them instead of 5 pure ones. For illustrative purposes, extraordinary care was taken to set *G03* input flags (*IOPs*) in such a way as to obtain the pure behaviour of a given acceleration technique. Of course, in most programs a mix of methods is invoked by default, which on average is more efficient than any given acceleration approach by itself.

The first example is a $(CH_3)_2CHOH$ calculation started from the extended Huckel guess (*guess = huckel*). Figure 1 compares the performance of the Roothaan algorithm, DIIS, EDIIS, and level shift with the parameter value chosen from several different runs, as well as two other parameters on either side of the optimal value. We note that all methods converge to the same minimum. The speed of convergence is quite slow for the Roothan method. The DIIS approach requires very few iterations ($\sim$10) to converge the energy to high accuracy. Interestingly enough, the best *damping* approach – EDIIS – and the level shift with the optimal parameter

---

[2]Reference [36] has made statements that the EDIIS mixing as defined by equation (35) is somehow unsuitable for DFT calculations. Unfortunately, this claimed "inferiority" of equation (35) in DFT was not actually demonstrated in any benchmarks. In our view, this criticism is irrelevant, and in practice equation (35) does work nicely with DFT methods. The supposedly "correct" mixing scheme from reference [36] assumes that the optimally interpolated density will necessarily be close to the last iterate. In contrast, EDIIS' equation (35) makes no such assumptions, and is fully symmetric with respect to any of the previous iterates. In situations where the EDIIS is most useful (i.e. early iterations with large changes in the density) the last iterate may not have the largest mixing coefficient, so this is why symmetric equation (35) that does not favour any of the previous iterates is actually prefered to *asymmetric* mixing formulae, such as the one in reference [36].
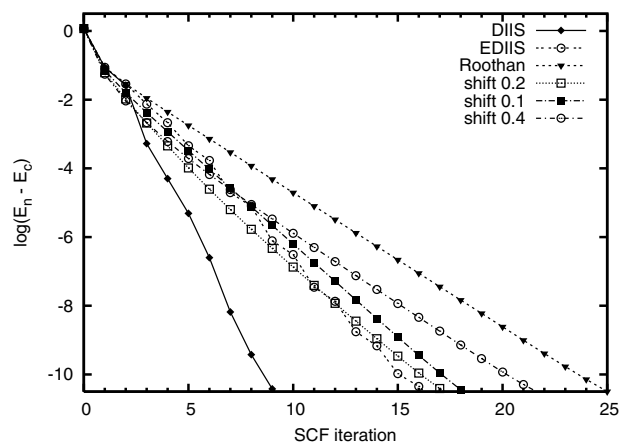
FIGURE 1. Comparison of SCF convergence patterns for $(CH_3)_2CHOH$ with the *Huckel* guess at the RHF/6-31G(d) level of theory. E_n is the current SCF iteration energy and E_c is the converged energy value of $-193.1\,150\,574\,013$ Hartrees.
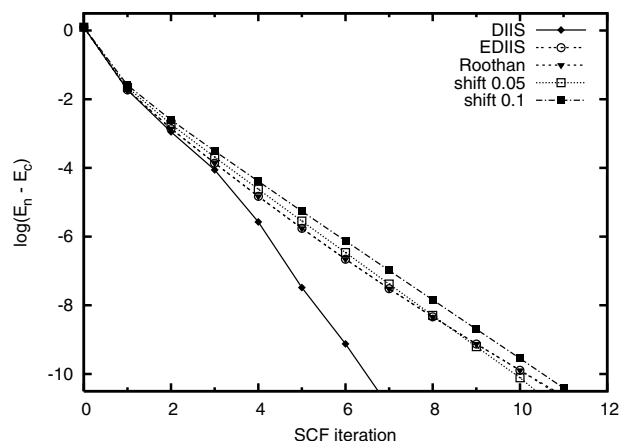

FIGURE 2. Comparison of SCF convergence patterns for C_6H_6 with the *Huckel* guess at the RHF/6-31G(d) level of theory. E_n is the current SCF iteration energy and E_c is the converged energy value of $-230.7\,031\,370\,193$ Hartrees.

reach the minimum at a similar speed. Inspection of the EDIIS mixing coefficients indicates that only one or two the most recent density matrices were used at any given iteration, yielding the observed wiggles in Figure 1.

The next example is benzene (C_6H_6) starting from the extended Huckel guess. It is worth noting that the Roothan algorithm converges rather quickly by itself (Fig. 2), perhaps due to the high symmetry of the system. DIIS improves upon that. The EDIIS curve coincides with the fixed point results (inspection of the coefficients shows that only the last density is used at every iteration). Finally, the optimal level shift is zero, and so any non-zero value makes the calculation slower. Here again, EDIIS and level shift with an optimal value (zero) yield the same convergence rate.

Now, to demonstrate a case where some methods do not converge, we run a calculation for the same iso-propyl-alcohol with a poor initial guess, specifically, the *core* guess obtained from diagonalizing the core Hamiltonian consisting of kinetic energy and nuclear attraction terms (h in Eq. (7)). We note that as a general rule, one can often create an example where the SCF does not converge by taking an otherwise well behaved system
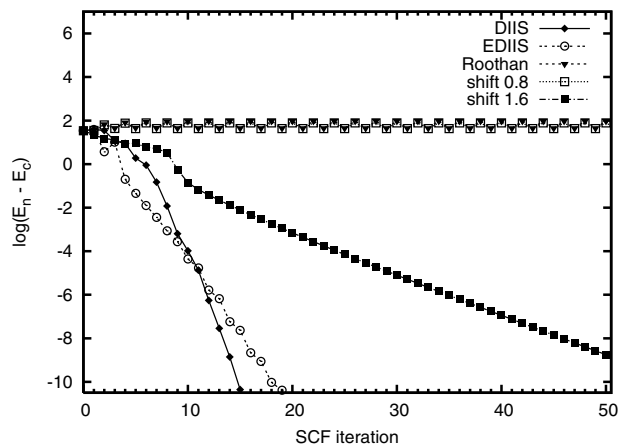
FIGURE 3. Comparison of SCF convergence patterns for $(CH_3)_2CHOH$ with the *core* guess at the RHF/6-31G(d) level of theory. E_n is the current SCF iteration energy and E_c is the converged energy value of –193.1 150 574 013 Hartrees.

and running it with a poor guess. In this example (Fig. 3), the fixed point iteration oscillates between 2 non-stationary densities and goes nowhere. DIIS does take a few iterations to get close to the minimum, but once nearby, it quickly finds it. EDIIS steadily lowers the energy in the initial iterations surpassing all other techniques. However, near convergence, EDIIS is outperformed by DIIS. The minimum value for level shift required to make this calculation converge is above 0.8 Hartree, and the value of 1.6 Hartree (presented in Fig. 3) does work as shown. Overall, such a large fixed level shift is neither particularly quick for the initial nor for the final convergence.

The last example, NiC at 1.8 Å, is a very challenging case. To make it even more interesting, we have used a poor initial guess (core guess) on purpose. Fixed point iteration does not converge (Fig. 4), and neither does DIIS. In contrast, EDIIS converges (albeit not particularly quickly compared to the earlier examples), and finds a lower energy stationary point compared to the level shift calculation. The level shift value needed to guarantee convergence is more than 1.6 Hartree. The convergent value of 3.2 Hartrees does help at early iterations to get near a minimum, however, at later stages such a large level shift leads to very slow convergence.

6. CONCLUSIONS AND FUTURE OUTLOOK

Among the approaches for SCF convergence acceleration discussed in this paper, DIIS stands out as the fastest method to get to a minimum once in the convergence region. Considering the observed rate of convergence (the drop in the energy error 10-fold per iteration) DIIS may be quite difficult to surpass except for quadratically convergent SCF with the exact Hessian. Note, however, that the quadratically convergent SCF requires the equivalent of several Fock matrix constructions per iteration, thus in terms of the total computational expense, it is much slower than DIIS by a wide margin. Quite surprising is that DIIS often does work even when initially far from convergence, as some examples indicate. Understanding the behaviour of DIIS from a rigorous mathematical point of view could be beneficial to develop further improvements to it. Another interesting question is that assuming that there exists just one proper *aufbau* solution for a given system, can it be shown that DIIS always converges in such a setting?

EDIIS is noticeable for its robustness, ease of implementation (similar to DIIS), and the broader exploration of the electronic space on its way to a stationary point. It is advisable to switch to the DIIS once a minimum is nearby, judged for example, by the maximum error in the commutator matrix $[\mathcal{F}_i, \mathcal{D}_i]$ [22]. In restricted HF and DFT methods where some subset of systems happen to have solutions with fractional occupations, EDIIS
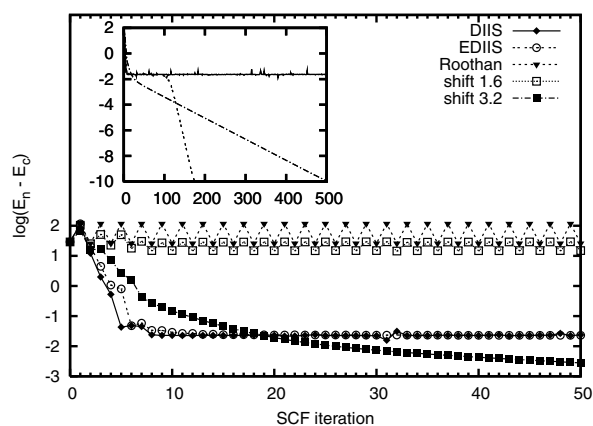
FIGURE 4. Comparison of SCF convergence patterns for NiC at 1.8 Å with multiple energy minima with the *core* guess at the RHF/6-31G(d) level of theory. E_n is the current SCF iteration energy and E_c is the converged energy. E_c(EDIIS) = −1544.0 527 099 012, E_c(level shift) = −1543.9 308 695 940 Hartrees. The inset describes the behaviour of DIIS, EDIIS and level shift for hundreds of SCF cycles where it becomes clear if a given method converges or not.

will flag them [5, 8, 22]. One needs to detect such cases, and either force a solution with integer occupations (and, possibly, *aufbau* violations), or alternatively, optimize a state with fractional occupations [8].

The level shift approach needs an adjustable parameter to make a given calculation converge. As we have shown in Section 5, the upper limit on the fixed level shift value guaranteeing convergence may be quite large, leading to slow convergence when near a minimum. Calculations with level shift also invariably gravitate toward a higher energy local minima when multiple minima are close in energy. We should note here that recent works employing a trust radius concept [13, 35, 36] do seem to make the level shift method more usable by aiming to ensure an energy reduction for the new density by iteratively adjusting the level shift value [35]. However, such a scheme is arguably more complex than DIIS or EDIIS, and does preserve the usually undesired tendency of the level shift method to converge to the closest local minimum [13].

With a range of acceleration techniques available that are either robust or quick to reach a local minimum, and which can be seamlessly combined, the only serious practical issue remaining is how to quickly explore the various local minima in multiple minima systems. For such a purpose, one could probably employ various stochastic and dynamics based approaches. For example, generalized simulated annealing has already been tried for RHF wavefunction optimization [10], and it would be interesting to see how such an approach performs for locating the global minimum.

From a more formal mathematical point of view, only few of the SCF iteration properties are fully characterized. The existence of a proper *aufbau* solution was originally proven for the *unrestricted* HF [23], and then both for the *unrestricted* HF and *extended* KS in a somewhat indirect way in the context of the relaxed constraints algorithms [5, 7]. The absolute convergence of the level shift has been discussed in references [4] and [13]. The works on the relaxed constraints algorithms helped to elucidate the origin of the energy oscillations in the Roothan algorithm, and also uncovered some of the peculiar properties of the HF energy functional [6, 7].

References

[1] G.B. Bacskay, A quadratically convergent Hartree-Fock (QC-SCF) method - application to closed shell systems. *Chem. Phys.* **61** (1981) 385–404.

[2] S.P. Bhattacharyya, Accelerated convergence in SCF calculations and level shifting technique. *Chem. Phys. Lett.* **56** (1978) 395–398.

[3] S.F. Boys, Electronic wave functions. 1. A general method of calculation for the stationary states of any molecular system. *Proc. R. Soc. Lond. A* **200** (1950) 542–554.

[4] E. Cancés, in *Mathematical Models and Methods for ab initio Quantum Chemistry*, M. Defranceschi and C. Le Bris Eds., *Lecture Notes in Chemistry* **74**, Springer, Berlin (2000).

[5] E. Cancés, Self-consistent field algorithms for Kohn-Sham models with fractional occupation numbers. *J. Chem. Phys.* **114** (2001) 10616–10622.

[6] E. Cancés and C. Le Bris, On the convergence of SCF algorithms for the Hartree-Fock equations. *ESAIM: M2AN* **34** (2000) 749–774.

[7] E. Cancés and C. Le Bris, Can we outperform the DIIS approach for electronic structure calculations? *Int. J. Quantum Chem.* **79** (2000) 82–90.

[8] E. Cancès, K.N. Kudin, G.E. Scuseria and G. Turinici, Quadratically convergent algorithm for fractional occupation numbers in density functional theory. *J. Chem. Phys.* **118** (2003) 5364–5368.

[9] A.D. Daniels and G.E. Scuseria, Converging difficult SCF cases with conjugate gradient density matrix search. *Phys. Chem. Chem. Phys.* **2** (2000) 2173–2176.

[10] M.D. de Andrade, K.C. Mundim and L.A.C. Malbouisson, GSA algorithm applied to electronic structure: Hartree-Fock-GSA method. *Int. J. Quant. Chem.* **103** (2005) 493–499.

[11] R.M. Dreizler and E.K.U. Gross, *Density functional theory*. Springer, Berlin (1990).

[12] A. Fouqueau, S. Mer, M.E. Casida, L.M.L. Daku, A. Hauser, T. Mineva and F. Neese, Comparison of density functionals for energy and structural differences between the high-[T-5(2g) : (t(2g))(4)(e(g))(2)] and low-[(1)A(1g) : (t(2g))(6)(e(g))(0)] spin states of the hexaquoferrous cation [Fe(H$_2$O)(6)](2$^+$). *J. Chem. Phys.* **120** (2004) 9473–9486.

[13] J.B. Francisco, J.M. Martinez and L. Martinez, Globally convergent trust-region methods for self-consitent field electronic structure calculations. *J. Chem. Phys.* **121** (2004) 10863–10878.

[14] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, and J.A. Pople, *Gaussian 03, Revision C.02.* Gaussian Inc., Wallingford CT (2004).

[15] D.R. Hartree, *The calculation of atomic structures*. Wiley (1957).

[16] T. Helgaker, H. Larsen, J. Olsen and P. Jorgensen, Direct optimization of the AO density matrix in Hartree-Fock and Kohn-Sham theories. *Chem. Phys. Lett.* **327** (2000) 397–403.

[17] P. Hohenberg and W. Kohn, Inhomogeneous electron gas. *Phys. Rev.* **136** (1964) 864B–B871.

[18] H. Hsu, E.R. Davidson and R.M. Pitzer, SCF method for hole states. *J. Chem. Phys.* **65** (1976) 609–613.

[19] B.G. Johnson, P.M.W. Gill and J.A. Pople, The performance of a family of Density Functional methods. *J. Chem. Phys.* **98** (1993) 5612–5626.

[20] G. Karlstrom, Dynamical damping based on energy minimization for use in ab initio molecular-orbital SCF calculations. *Chem. Phys. Lett.* **67** (1979) 348–350.

[21] W. Kohn and L.J. Sham, Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140** (1965) A1133–A1138.

[22] K.N. Kudin, G.E. Scuseria and E. Cancès, A black-box self-consistent field convergence algorithm: One step closer. *J. Chem. Phys.* **116** (2002) 8255–8261.

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

[24] A.V. Mitin, The dynamic level shift method for improving the convergence of the SCF procdure. *J. Comput. Chem.* **9** (1988) 107–110.

[25] M.A. Natiello and G.E. Scuseria, Convergence properties of Hartree-Fock SCF molecular calculations. *Int. J. Quantum Chem.* **26** (1984) 1039–1049.

[26] P. Pulay, Convergence acceleration of iterative sequences - the case of SCF iteration. *Chem. Phys. Lett.* **73** (1980) 393–398.

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

[28] A.D. Rabuck and G.E. Scuseria, Improving self-consistent field convergence by varying occupation numbers. *J. Chem. Phys.* **110** (1999) 695–700.

[29] M. Ramek and H.P. Fritzer, New aspects of dynamical damping in ab initio molecular SCF calculations. *Comput. Chem.* **6** (1982) 165–168.

[30] C.C.J. Roothan, New developments in molecular orbital theory. *Rev. Mod. Phys.* **23** (1951) 69–89.

[31] Y. Saad and M.H. Schultz, GMRES - a generalized minimal residual algorithm for solving nonsymmetric linear-systems. *SIAM J. Sci. Stat. Comput.* **7** (1986) 856–869.

[32] V.R. Saunders and I.H. Hillier, Level-shifting method for converging closed shell Hartree-Fock wave-functions. *Int. J. Quant. Chem.* **7** (1973) 699–705.

[33] N.E. Schultz, Y. Zhao and D.G. Truhlar, Databases for transition element bonding: Metal-metal bond energies and bond lengths and their use to test hybrid, hybrid meta, and meta density functionals and generalized gradient approximations. *J. Phys. Chem. A* **109** (2005) 4388–4403.

[34] G.E. Scuseria, Linear scaling density functional calculations with Gaussian orbitals. *J. Phys. Chem. A* **103** (1999) 4782–4790.

[35] L. Thogersen, J. Olsen, D. Yeager, P. Jorgensen, P. Salek and T. Helgaker, The trust-region self-consistent field method: Towards a black-box optimization in Hartree-Fock and Kohn-Sham theories. *J. Chem. Phys.* **121** (2004) 16–27.

[36] L. Thogersen, J. Olsen, A. Kohn, P. Jorgensen, P. Salek and T. Helgaker, The trust-region self-consitent field method in Kohn-Sham density-functional theory. *J. Chem. Phys.* **123** (2005) 074103.

[37] T.V. Voorhis and M. Head-Gordon, A geometric approach to direct minimization. *Mol. Phys.* **100** (2002) 1713–1721.

[38] O.V. Yazyev, K.N. Kudin and G.E. Scuseria, Efficient algorithm for band connectivity resolution. *Phys. Rev. B* **65** (2002) 205117.

[39] M.C. Zerner and M. Hehenberger, Dynamical damping scheme for converging molecular SCF calculations. *Chem. Phys. Lett.* **62** (1979) 550–554.