

Do You Have SCF Stability and Convergence Problems?

H.B. Schlegel
Department of Chemistry
Wayne State University
Detroit, Michigan 48202, USA.

J.J.W. McDouall
Department of Chemistry
University of Manchester
Manchester M13 9PL, U.K.

ABSTRACT Because the Hartree-Fock (HF) equations are non-linear in the molecular orbitals (MO's), they must be solved iteratively to obtain a self-consistent set of MO's (i.e. the SCF method). When iterations of the SCF process converge, the energy is stationary with respect to infinitesimal variations in the orbitals. Whether this stationary point is stable (a local minimum) or unstable (a maximum in one or more directions) depends on the second derivatives of the energy with respect to the orbital variations. The internal and external stabilities or instabilities of various constrained HF wavefunctions (real and complex, restricted, unrestricted and general HF) are discussed. A simple iterative scheme does not always lead to convergence of the SCF equations. The connection between SCF stability and convergence is examined and a number of methods for improving SCF convergence are discussed, including damping, level shifting, extrapolation, univariate search, direct inversion of iterative subspace (DIIS) and quadratically convergent SCF. Techniques for constructing better initial guess MO's are also considered.

1. Introduction.

Most quantum chemical calculations using molecular orbital methods determine the orbitals by some iterative method. Typically, for each electron the orbital is optimized in the field of all the other electrons. This process is repeated until convergence is achieved and each orbital is optimal in the self consistent field (SCF) of all the other electrons. Hartree-Fock theory is the result when this approach is used with a single determinantal wavefunction; alternatively, this approach applied to a multi-configurational wavefunction yields an MC-SCF calculation. Because of the iterative nature of the SCF procedure, convergence problems are sometimes encountered. Furthermore, when convergence is achieved, there is often no guarantee that the solution to the SCF problem is stable with respect to all permitted changes in the wavefunction.

In this Chapter we will address some of the issues related to SCF convergence and stability within the single determinantal, Hartree-Fock (HF) picture. Problems with stability and convergence in MCSCF calculations are beyond the scope of this Chapter. Section 2 briefly outlines the Hartree-Fock approach for molecular orbitals expressed as linear combinations of atomic orbitals (LCAO-MO), i. e. the Roothaan-Hall equations. The question of the stability of the SCF wavefunction under various constraints is discussed in Section 3. The next Section deals with algorithms for obtaining SCF convergence. Section 5 covers various methods for obtaining initial estimates of the molecular orbitals to start the SCF process.

2. Hartree-Fock Theory and SCF Calculations

In the Hartree-Fock model, the electronic wavefunction for an n -electron system is represented by a single Slater determinant, $|\Psi_0\rangle$, of n spin orbitals $\{\Phi_i\}$

$$(1) \quad |\Psi_0\rangle = |\Phi_1 \Phi_2 \cdots \Phi_n\rangle$$

The orbitals are optimized according to the variation principle such that the energy functional

$$(2) \quad E_0 = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle / \langle \Psi_0 | \Psi_0 \rangle$$

is minimized. In eq 2, \hat{H} , represents the full electronic hamiltonian. The derivation of the Hartree-Fock equation, which determines the optimal spin orbitals, may be found in several standard texts¹⁻³ and will not be given here. We shall simply note that the Hartree-Fock equation is an eigenvalue equation of the form

$$(3) \quad \hat{F}(i) \Phi(\tau_i) = \epsilon \Phi(\tau_i)$$

where $\hat{F}(i)$ is an effective one-electron operator (for the i^{th} electron) called the Fock operator and is given by

$$(4) \quad \hat{F}(i) = \hat{h}(i) + \hat{V}^{\text{HF}}(i)$$

$$\text{where } \hat{h}(i) = -\frac{1}{2} \nabla_i^2 - \sum_A^{\text{nuclei}} \frac{Z_A}{r_{iA}}, \quad \hat{V}^{\text{HF}}(i) = \sum_j \hat{J}_j(i) - \hat{K}_j(i)$$

$$\text{and } \hat{J}_j(1) \Phi_i(1) = \int \Phi_j(2) \frac{1}{r_{12}} \Phi_j(2) \Phi_i(1) d\tau_2$$

$$\hat{K}_j(1) \Phi_i(1) = \int \Phi_j(2) \frac{1}{r_{12}} \Phi_i(2) \Phi_j(1) d\tau_2,$$

in which, ∇_i^2 is the Laplacian kinetic energy operator for electron i , Z_A is the charge on nucleus A , r_{iA} is the distance between electron i and nucleus A and $\bar{V}^{\text{HF}}(i)$ is the average potential experienced by electron i due to all other electrons. Hence, eq 3 replaces the instantaneous inter-electron repulsion by an averaged field. This is the central idea in the Hartree-Fock approximation.

To cast eq 3 into a form suitable for computation, we write the spin orbitals $\{\Phi_i\}$ as a linear combination of N one-electron basis functions $\{\chi_v\}$

$$(5) \quad \Phi_i = \sum_v^N C_{vi} \chi_v$$

This approach was first applied to the case of closed-shell systems by Roothaan and Hall⁴ and is commonly referred to as the restricted Hartree-Fock (RHF) method because all doubly occupied orbitals are constrained to have the same spatial form. The method was later extended to open-shell systems by Pople and Nesbet⁵, who incorporated the additional flexibility of allowing a different spatial form for electrons of α and β spin. This method is known as the unrestricted Hartree-Fock (UHF) approach. These ideas are expanded further in Section 3; for the moment, no distinction is made between RHF and UHF determinants. Substituting eq 5 into eq 3 yields

$$(6) \quad \sum_v^N \hat{F} \chi_v C_{vi} = \epsilon_i \sum_v^N \chi_v C_{vi}$$

premultiplying eq 6 by χ_μ and integrating gives

$$(7) \quad \sum_v^n \{F_{\mu v} - \epsilon_i S_{\mu v}\} C_{vi} = 0 \quad i=1,2 \dots N$$

where $F_{\mu v} = \langle \chi_\mu | \hat{F} | \chi_v \rangle$ and $S_{\mu v} = \langle \chi_\mu | \chi_v \rangle$

The Fock matrix, in turn can be written in terms of the density matrix, \mathbf{P} , and the one and two electron integrals ($h_{\mu v}$ and $(\mu\lambda||\nu\sigma)$, respectively).

$$(8) \quad F_{\mu v} = h_{\mu v} + \sum_{\lambda\sigma} (\mu\lambda||\nu\sigma) P_{\lambda\sigma}$$

$$(9) \quad P_{\lambda\sigma} = \sum_i C_{\lambda i}^* C_{\sigma i}, \quad h_{\mu v} = \int \chi_\mu(1)^* \hat{h}(1) \chi_v(1) d\tau_1$$

$$(\mu\lambda||\nu\sigma) = \int \chi_\mu(1)^* \chi_\lambda(2)^* \frac{1}{r_{12}} (\chi_\nu(1) \chi_\sigma(2) - \chi_\sigma(1) \chi_\nu(2)) d\tau_1 d\tau_2$$

Thus the basis set expansion has turned the integro-differential Hartree-Fock equation into a matrix equation,

$$(10) \quad \mathbf{F} \mathbf{C} = \mathbf{S} \mathbf{C} \boldsymbol{\epsilon}$$

where \mathbf{F} is the matrix representation of the Fock operator, \mathbf{C} is the matrix of molecular orbital coefficients, $\{C_{vi}\}$ (eq 5), \mathbf{S} is the overlap matrix of basis functions and $\boldsymbol{\epsilon}$ is a diagonal matrix containing the one-electron orbital energies. Equation 10 is nonlinear in the MO coefficients, since \hat{J} and \hat{K} in \hat{F} depend on \mathbf{C} . It is therefore necessary to solve eq 10 iteratively. The Fock matrix is built using an initial guess of the MO coefficients, $\{C_{vi}\}$. Equation 10 is solved for a new set of $\{C_{vi}\}$ which are then used to recompute the Fock matrix. This process is iterated until the change in the molecular orbitals and the orbital energies falls below some chosen threshold. The orbitals obtained in this manner are said to be self-consistent and the average potential that they describe for each electron is termed the self-consistent field (SCF).

Because of the nonlinearity of eq 10, the simple iterative procedure described above does not always converge. It may oscillate or even diverge, depending on the initial guess and the nature of the electronic structure being calculated. Even when convergence is obtained, eq 10 simply ensures that the energy is stationary with respect to the first-order variation of the $\{C_{vi}\}$. To determine the nature of the stationary point (i.e. whether the energy is a maximum or a minimum), the second-order variation of the energy with respect to the $\{C_{vi}\}$ must be examined⁶. This is the subject of the next section.

3. SCF Constraints and Stability.

Each member of the set of spin orbitals $\{\Phi_i\}$ is a function of the coordinates (τ) of a single electron. The electronic coordinate, τ , may be separated into a spatial component, which typically depends on the Cartesian coordinates, \vec{r} , of the electron and a spin coordinate, σ . In solving the SCF equations, it is common practice to impose certain constraints on the nature of the spin orbitals and carry out the variation of Ψ_0 in a restricted space. Typical examples of constraints include representing each Φ_i as a product of a function of Cartesian coordinates and a function of spin coordinates and restricting the functions to be real. Hence, to establish that the solution to eq 10 (within the constraints imposed) does lead to a local minimum in the energy (eq 2), it is necessary to examine its second-order variation and show that it is positive. Once a constrained minimum has been located, the consequences of removing one or more of the constraints can be investigated. The possible outcomes of removing constraints on Ψ_0 are: (a) $E(\Psi_0)$ remains at a local minimum, or (b) the first-order variation of Ψ_0 is zero but $E(\Psi_0)$ is no longer at local minimum, or (c) $E(\Psi_0)$ is no longer stationary. The stability conditions were originally derived by Thouless⁶; Cizek and Paldus⁷ presented the stability problem from a chemical point of view. Systematic discussion of constraints and stability conditions were presented later by Ostlund⁸, Fukutome⁹ and Seeger and Pople¹⁰. The latter approach¹⁰ will be followed in the next section.

General Hartree-Fock (GHF) In general, the spin orbital $\Phi(r,\sigma)$ is a complex function, usually taken to be part of an orthonormal set, i.e.

$$(11) \quad \int \Phi_i^* \Phi_j \, d\tau = \delta_{ij}$$

Expanding Φ_i in spin space over the two spin basis functions $\alpha(\sigma)$ and $\beta(\sigma)$ gives

$$(12) \quad \Phi_i(r, \sigma) = \phi_i^\alpha(r) \alpha(\sigma) + \phi_i^\beta(r) \beta(\sigma)$$

where ϕ_i^α and ϕ_i^β are functions of the Cartesian coordinates alone (spatial orbitals). In general ϕ_i^α and ϕ_i^β are complex and may be separated into their real and imaginary components,

$$(13) \quad \Phi_i^\alpha = \text{Re}(\phi_i^\alpha) + i \text{Im}(\phi_i^\alpha), \quad \Phi_i^\beta = \text{Re}(\phi_i^\beta) + i \text{Im}(\phi_i^\beta)$$

which leads to the following general form for the spin orbitals in terms of the N spatial basis functions $\{\chi\}$

$$(14) \quad \Phi_i(r, \sigma) = \sum_{\mu}^N [C_{\mu i}^{\alpha} \chi_{\mu}(r) \alpha(\sigma) + C_{\mu i}^{\beta} \chi_{\mu}(r) \beta(\sigma)] \quad (\text{GHF orbitals})$$

where $C_{\mu i}^{\alpha}$ and $C_{\mu i}^{\beta}$ are complex. The use of the full expression in eq 14, defining the spin orbitals in terms of $4N$ real numbers (where N is the number of spatial basis functions), leads to what has been termed general Hartree-Fock theory (GHF). Note that the GHF wavefunctions are not eigenfunctions of the spin operators \hat{S}^2 and \hat{S}_z . The spatial functions $\{\chi\}$ and the MO coefficients can be constrained to be real, yielding the real-GHF approximation. Other levels of Hartree-Fock theory can be obtained by applying other constraints to eq 14.

Unrestricted Hartree-Fock (UHF) The most commonly applied constraint is to write each spin orbital as a product of the form $\chi_{\mu}(r) \alpha(\sigma)$ or $\chi_{\mu}(r) \beta(\sigma)$ rather than the general linear combination of eq 14 (i.e. the UHF spin orbitals are either pure α or pure β rather than mixed). The α and β orbitals are independent complex functions and this approximation is usually referred to as unrestricted Hartree-Fock theory (UHF)

$$(15) \quad \Phi_i(r, \alpha) = \sum_{\mu} C_{\mu i}^{\alpha} \chi_{\mu}(r) \alpha(\sigma) \quad \text{and} \quad \Phi_i(r, \beta) = \sum_{\mu} C_{\mu i}^{\beta} \chi_{\mu}(r) \beta(\sigma) \quad (\text{UHF orbitals})$$

The UHF wavefunctions are eigenfunctions of the \hat{S}_z but not of \hat{S}^2 , the total spin operator¹¹. At this level, each spin orbital may be specified by N complex coefficients $C_{\mu i}^{\alpha}$ or $C_{\mu i}^{\beta}$. The additional constraint of requiring the spatial functions $\{\chi\}$ and the MO coefficients to be real leads to the real-UHF approximation.

Restricted Hartree-Fock (RHF) The next constraint relates the spatial parts of orbitals of different spin. If the numbers of occupied α and β orbitals are n_{α} and n_{β} , where $n_{\alpha} \geq n_{\beta}$, then n_{β} of the α orbitals may be restricted to have the same spatial functions as the β orbitals. These n_{β} orbitals are then "doubly occupied" and the remaining $(n_{\alpha} - n_{\beta})$ α orbitals are "singly occupied"; this approximation is referred to as restricted Hartree-Fock theory (RHF). All doubly occupied orbitals (i.e. closed-shells with $n_{\alpha} = n_{\beta}$) is an important special case in the RHF scheme.

$$(16) \quad \Phi_i(r, \sigma) = \sum_{\mu} C_{\mu i} \chi_{\mu}(r) \alpha(\sigma) \quad \text{and} \quad \bar{\Phi}_i(r, \sigma) = \sum_{\mu} C_{\mu i} \chi_{\mu}(r) \beta(\sigma) \quad (\text{RHF orbitals})$$

All RHF wavefunctions (whether closed- or open-shell) have the advantage that they are eigenfunctions of the spin operators \hat{S}^2 and \hat{S}_z , and so represent pure spin states. For the examination of the stability of RHF wavefunctions, only closed shell RHF will be considered, since open shell RHF (ROHF) wavefunctions are always unstable with respect UHF.

Stability of HF Wavefunctions A Hartree-Fock wavefunction is considered stable if none of the possible changes in the one electron orbitals leads to a lower energy. It can be shown readily that the energy is invariant with respect to mixing the occupied orbitals in a set with themselves (e.g. doubly occupied with doubly occupied for RHF, α with α for UHF, etc.). In addition to the n occupied orbitals, there are $N-n$ virtual orbitals ($\Phi_{n+1} \cdots \Phi_N$) which arise from the choice of the one-electron basis functions in eq 5. Thus, it is necessary to examine how the energy changes when these virtual orbitals are allowed to mix with occupied orbitals¹³. Let an occupied orbital change according to:

$$(17) \quad \Phi_i \rightarrow \Phi_i' = \Phi_i + \sum_a D_{ia} \Phi_a$$

where the sum extends over all virtual orbitals. The standard notation of indices i, j, k , etc. for occupied orbitals, indices a, b, c , etc. for virtual orbitals and indices p, q, r , etc. for any orbital (occupied or virtual) is used throughout. When the occupied orbitals Φ_i in eq 1 are replaced by Φ_i' and the wavefunction is expanded, the result can be written in terms of the Hartree-Fock ground state determinant and a sum over excited determinants obtained by replacing all possible combinations of occupied spin orbitals Φ_i ($1 \leq i \leq n$) with all possible combinations of virtual spin orbitals Φ_a ($n+1 \leq a \leq N$):

$$(18) \quad |\Psi\rangle = D_0 |\Psi_0\rangle + \sum_{ia} D_{ia} |\Psi_i^a\rangle + \frac{1}{2!} \sum_{ij ab} D_{ia} D_{jb} |\Psi_{ij}^{ab}\rangle + \frac{1}{3!} \sum_{ijk abc} D_{ia} D_{jb} D_{ck} |\Psi_{ijk}^{abc}\rangle + \cdots$$

where

$$(19) \quad |\Psi_0\rangle = |\cdots \Phi_i \Phi_j \Phi_k \cdots\rangle, \quad |\Psi_i^a\rangle = |\cdots \Phi_a \Phi_j \Phi_k \cdots\rangle, \quad |\Psi_{ij}^{ab}\rangle = |\cdots \Phi_a \Phi_b \Phi_k \cdots\rangle, \quad \text{etc.}$$

and D_0 is introduced to allow for renormalization. Likewise, the energy, eq 2, can be expanded:

$$(20) \quad E = \langle \Psi | \hat{H} | \Psi \rangle / \langle \Psi | \Psi \rangle$$

$$= \left\{ D_0^* D_0 E_0 + \sum_{ia} [D_0^* D_{ia} \langle \Psi_0 | \hat{H} | \Psi_i^a \rangle + D_{ia}^* D_0 \langle \Psi_i^a | \hat{H} | \Psi_0 \rangle] \right.$$

$$+ \sum_{ij ab} \left[\frac{1}{2} D_0^* D_{ia} D_{jb} \langle \Psi_0 | \hat{H} | \Psi_{ij}^{ab} \rangle + \frac{1}{2} D_{ia}^* D_{jb}^* D_0 \langle \Psi_{ij}^{ab} | \hat{H} | \Psi_0 \rangle \right.$$

$$\left. \left. + D_{ia}^* D_{jb} \langle \Psi_i^a | \hat{H} | \Psi_j^b \rangle \right] + \cdots \right\} (D_0^* D_0 + \sum_{kc} D_{kc}^* D_{kc})^{-1}$$

For the stability analysis, it is sufficient to adopt intermediate normalization, i.e. $D_0 = 1$. Since Ψ_0 is a converged Hartree-Fock wavefunction, the matrix elements involving the ground state and single excitations, e.g. $\langle \Psi_0 | \hat{H} | \Psi_i^a \rangle$, are zero by Brillouin's theorem. If the expansion is truncated at second order, the energy can be written in a convenient matrix form⁷:

$$\begin{aligned}
 (21) \quad E &= E_0 + \frac{1}{2} \sum_{ij\ ab} [D_{ia} D_{jb} \langle \Psi_0 | \hat{H} | \Psi_{ij}^{ab} \rangle + D_{ia}^* D_{jb}^* \langle \Psi_{ij}^{ab} | \hat{H} | \Psi_0 \rangle \\
 &\quad + 2 D_{ia}^* D_{jb} \langle \Psi_i^a | \hat{H} - E_0 | \Psi_j^b \rangle] \\
 &= E_0 + \frac{1}{2} \begin{bmatrix} \mathbf{D} \\ \mathbf{D}^* \end{bmatrix}^\dagger \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{bmatrix} \begin{bmatrix} \mathbf{D} \\ \mathbf{D}^* \end{bmatrix} = E_0 + \frac{1}{2} \tilde{\mathbf{D}}^\dagger \mathbf{H} \tilde{\mathbf{D}}
 \end{aligned}$$

where \mathbf{D}^* and \mathbf{D}^\dagger are the complex and hermitian conjugates of \mathbf{D} , the matrix of the coefficients D_{ia} . The matrix \mathbf{H} is composed of submatrices \mathbf{A} and \mathbf{B} , which are given by

$$(22) \quad A_{ia\ jb} = \langle \Psi_i^a | \hat{H} - E_0 | \Psi_j^b \rangle = (\epsilon_a - \epsilon_j) \delta_{ab} \delta_{ij} + (aj||ib),$$

$$B_{ia\ jb} = \langle \Psi_{ij}^{ab} | \hat{H} | \Psi_0 \rangle = (ab||ij)$$

Note that \mathbf{A} is hermitian and \mathbf{B} is symmetric. For E to be a minimum with respect to the variations $\tilde{\mathbf{D}}$, the matrix \mathbf{H} must be positive semidefinite, i.e. no negative eigenvalues. For open shell systems, a zero eigenvalue is permissible, since the direction of spin quantization is indeterminate. For general Hartree-Fock theory, all variations in $\tilde{\mathbf{D}}$ are permitted. For the other, constrained HF wavefunctions, only certain selected variations are allowed and hence only specific portions of \mathbf{H} need to be positive semidefinite to ensure stability.

Stability of Constrained Hartree-Fock Wavefunctions Figure 1 summarizes the types of stabilities that will be considered. A wavefunction can be tested for stability within the constraint (internal stability) or for stability with respect to a less constrained wavefunction (external stability). In the Figure, releasing constraints corresponds to moving down and/or to the right.

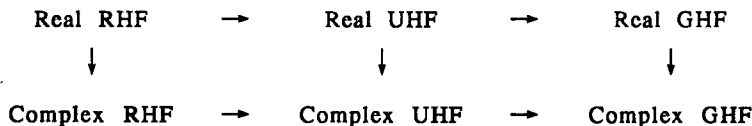


Figure 1. Hartree-Fock wavefunctions and the types stability considered.

One constraint that can be applied to RHF, UHF and GHF wavefunctions is that the wavefunction remain real. In this case, the matrices \mathbf{A} and \mathbf{B} in eq 21 are real and the energy can be written as:

$$(23) \quad E = E_0 + \text{Re}(\mathbf{D}^\dagger) (\mathbf{A} + \mathbf{B}) \text{Re}(\mathbf{D}) + \text{Im}(\mathbf{D}^\dagger) (\mathbf{A} - \mathbf{B}) \text{Im}(\mathbf{D})$$

where $\text{Re}(\mathbf{D}) = 1/2 (\mathbf{D} + \mathbf{D}^*)$ and $\text{Im}(\mathbf{D}) = 1/2 (\mathbf{D} - \mathbf{D}^*)$. For internal stability, \mathbf{D} must be real and the matrix $\mathbf{A} + \mathbf{B}$ must be positive semidefinite. A non-zero imaginary component in \mathbf{D} will result in a complex wavefunction. Thus if $\mathbf{A} - \mathbf{B}$ has a negative eigenvalue, the energy will be lowered by making the wavefunction complex, i.e. the wavefunction is not stable with respect to becoming complex.

In a UHF wavefunction, the α spin orbitals cannot be mixed with the β spin orbitals. The excitations in \mathbf{D} can be divided into two groups: spin-conserving $\alpha \rightarrow \alpha$ and $\beta \rightarrow \beta$ excitations, \mathbf{D}' , and spin-unconserving $\alpha \rightarrow \beta$ and $\beta \rightarrow \alpha$ excitations, \mathbf{D}'' . Similarly, the matrices \mathbf{A} , \mathbf{B} and \mathbf{H} can be divided into spin-conserving and spin-unconserving components, designated by ' and ', respectively¹⁴. Only spin conserving excitations, \mathbf{D}' , maintain the constraints on a UHF wavefunction; hence, \mathbf{H}' must be positive semidefinite for a UHF wavefunction to be internally stable. Spin-unconserving excitations, \mathbf{D}'' , mix the α and β spin orbitals, changing the UHF wavefunction into a GHF wavefunction. Thus, if \mathbf{H}'' is not positive semidefinite, then the UHF wavefunction is GHF unstable. For UHF wavefunctions constrained to be real, $\mathbf{A}' + \mathbf{B}'$ and $\mathbf{A}'' + \mathbf{B}''$ have to be positive semidefinite for internal real UHF and external real GHF stability, respectively. Real UHF versus complex UHF stability is determined by $\mathbf{A}' - \mathbf{B}'$.

For a closed shell RHF wavefunction, all of the α spin orbitals must remain identical to the β spin orbitals. The excitations in \mathbf{D}' can be separated into 2 groups: $\mathbf{D}^{\alpha\alpha}$, the $\alpha \rightarrow \alpha$ excitations, and $\mathbf{D}^{\beta\beta}$, the $\beta \rightarrow \beta$ excitations in \mathbf{D}' . A linear combination of these 2 terms yield a singlet component ${}^1\mathbf{D}' = 1/\sqrt{2} (\mathbf{D}^{\alpha\alpha} + \mathbf{D}^{\beta\beta})$ for which the $\alpha \rightarrow \alpha$ and $\beta \rightarrow \beta$ excitations are the same and a triplet component ${}^3\mathbf{D}' = 1/\sqrt{2} (\mathbf{D}^{\alpha\alpha} - \mathbf{D}^{\beta\beta})$ for which the $\alpha \rightarrow \alpha$ and $\beta \rightarrow \beta$ excitations are opposite in sign; the singlet excitations maintain the constraints of the RHF wavefunction and the triplet excitations change the RHF wavefunction into a UHF wavefunction. Similarly, the matrices \mathbf{A}' , \mathbf{B}' and \mathbf{H}' can be divided into singlet and triplet components. Thus, for RHF internal stability, ${}^1\mathbf{H}'$ must be positive semidefinite; likewise, ${}^3\mathbf{H}'$ not positive semidefinite signals an instability in changing the RHF wavefunction to UHF. For a real RHF wavefunction, internal stability is determined by ${}^1\mathbf{A}' + {}^1\mathbf{B}'$, external stability with respect to a complex RHF wavefunction is determined by ${}^1\mathbf{A}' - {}^1\mathbf{B}'$.

The conditions for stability of the various Hartree-Fock wavefunctions are summarized in Table 1. The explicit formulas for the different \mathbf{A} 's, \mathbf{B} 's and \mathbf{H} 's are given in references 10.

Table 2 summarizes the results for nitrogen molecule, showing the various forms of instability which occur as the bond is stretched. The ground state electronic configuration of N_2 is shown in Figure 2. At bond lengths less than the equilibrium value, no instabilities occur; hence, all types of wavefunctions (RHF, UHF, GHF) lead to the same

Table 1. Matrices required to be positive semidefinite in order for an SCF solution of type **X** to be stable in a space of type **Y**

Y =	real RHF	complex RHF	real UHF	complex UHF	real GHF	complex GHF
X =						
real RHF	${}^1A'+{}^1B'$	${}^1A'-{}^1B'$	${}^3A'+{}^3B'$			
complex RHF		${}^1H'$		${}^3H'$		
real UHF			$A'+B'$	$A'-B'$	$A''+B''$	
complex UHF				H'		H''
real GHF					$A+B$	$A-B$
complex GHF						H

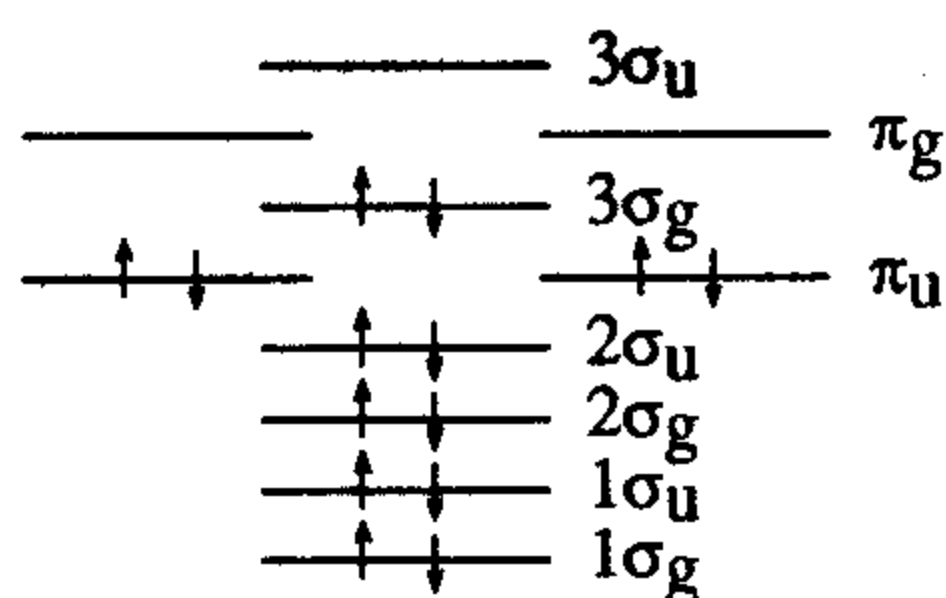


Figure 2. Electronic configuration of N_2 near R_{eq} .

Table 2. Total energies and instabilities that occur as the bond in N_2 is stretched.

R	real RHF	Types of instability	real UHF	Types of instability.
0.8	-106.680804	none	-106.680804	none
1.0	-107.419532	none	-107.419532	none
1.2	-107.487783	real UHF	-107.501203	GHF
1.4	-107.357815	complex RHF real UHF	-107.396565 ^b	internal UHF GHF
1.6	-107.184846 ^c	internal RHF complex RHF	-107.287416 ^d	internal UHF complex UHF, GHF

^a calculated with the STO-3G basis, $R_{eq} = 1.1339 \text{ \AA}$, total energies in hartree.

^b $E = -107.412014$ after re-optimization for internal UHF instability.

^c $E = -107.225669$ after re-optimization for internal RHF instability.

^d $E = -107.445187$ after re-optimization for internal UHF instability.

energy. As the bond is stretched, various instabilities arise from the mixing of the occupied π_u and the virtual π_g orbitals. At $R = 1.2 \text{ \AA}$, the RHF wavefunction is UHF unstable; the anti-bonding π_g mixes with the bonding π_u to give a lower energy real UHF wavefunction. In turn, the UHF wavefunction is GHF unstable. A complex instability sets in near $R = 1.4 \text{ \AA}$ for the RHF case and near 1.6 \AA for the UHF case. Although the initial UHF wavefunctions obtained for $R = 1.4$ and 1.6 \AA have lower energies lower than their RHF counterparts, the UHF solution themselves are internally unstable and can be optimized to give lower energies.

4. SCF Convergence

In general, the Hartree-Fock equations must be solved iteratively. In a simple SCF procedure, one cycle of the iteration consists of (a) constructing a new Fock matrix, F^{k+1} , using the current approximation to the MO coefficients, C^k , or the density matrix, P^k , and (b) obtaining the new (hopefully, improved) coefficient matrix C^{k+1} , by diagonalizing F^{k+1} .

$$(24) \quad F_{\mu\nu}^{k+1} = h_{\mu\nu} + \sum_{\lambda\sigma} (\mu\lambda|\nu\sigma) P_{\lambda\sigma}^k; \quad P_{\lambda\sigma}^k = \sum_j C_{\lambda j}^{k*} C_{\sigma j}^k$$

$$(25) \quad \sum_{\nu} F_{\mu\nu}^{k+1} C_{\nu p}^{k+1} = \sum_{\nu} S_{\mu\nu} C_{\nu p}^{k+1} \epsilon_p$$

This cycle is repeated until the changes in C or P are below a pre-set threshold. Unfortunately, this simple iteration scheme does not always converge; furthermore, the convergence difficulties are often more severe for open shell systems than for closed shells. A number of techniques for controlling and accelerating the convergence of SCF calculations are discussed in this section.

Damping One of the simplest and earliest used methods for forcing SCF convergence is damping¹⁶, whereby a weighted average of the old and the new matrices is used,

$$(26) \quad M^{k+1}' = (1-\alpha) M^{k+1} + \alpha M^k \quad 0 \leq \alpha < 1$$

where α is a constant. This can be applied to either the MO coefficient matrix, the density matrix or the Fock matrix (for MO coefficients, some care must be taken with phasing and mixing of degenerate orbitals). Damping has the effect of reducing the size of the step in the iteration and limits the tendency for oscillation. Typically a fixed value of α is chosen until the SCF has converged partially and $\alpha = 0$ (no damping) is used until final convergence is reached (the case where α is optimized is treated below under Univariate Search). Alternatively the damping factor can be adjusted dynamically¹⁷⁻¹⁹. One of the drawbacks of damping is that it does not preserve the orthonormality of the MO's or the idempotency of the density matrix²⁰. For the early stages of the SCF process, this is not a serious problem since the later (undamped) steps properly maintain the orthonormality of the MO's. Secondly, damping slows down the rate of convergence substantially;

however, in the absence of better techniques, slow convergence is better than no convergence.

Level Shifting²¹ This is an alternative to damping in controlling the size of the step taken in an SCF cycle. The energies of the virtual orbitals are increased by a constant amount (i.e. a constant is added to the diagonal elements of the virtual-virtual block of the Fock matrix in the MO basis). The effect of this can be seen by using perturbation theory to estimate the changes in the MO coefficients for an SCF cycle. Parallel to eq 17, let the new coefficients for the occupied orbitals, C_{vi}^{k+1} , be written in terms of the old coefficients plus a contribution from the virtual orbitals.

$$(27) \quad C_{vi}^{k+1} = C_{vi}^k + \sum_a D_{ia}^{k+1} C_{va}^k$$

Insertion of eq 27 into eq 25 and pre-multiplying by $C_{\mu a}^{k*}$ leads to

$$(28) \quad D_{ia}^{k+1} = - \sum_{\mu\nu} C_{\mu a}^{k*} F_{\mu\nu}^k C_{vi}^k / (\epsilon_a - \epsilon_i) = - F_{ai}^k / (\epsilon_a - \epsilon_i)$$

This estimate of the change in the MO coefficients is appropriate for the simple SCF procedure. If level shifting is used, ϵ_a is replaced by $\epsilon_a + b$ where $b > 0$; consequently, the changes in the MO coefficients are smaller than without level shifting, and the tendency for the SCF to oscillate is reduced. The level shift can be fixed or adjusted dynamically^{21,22}. While damping reduces the step by the same amount for each orbital, level shifting reduces the occupied-virtual mixing more for those pairs with small energy differences.

Extrapolation Just as damping can stabilize the early phase of an SCF calculation, extrapolation can speed up the final convergence. The Aitken extrapolation²³ assumes the matrix elements form a geometric series (e.g. $(M^{k+1} - M^\infty) / (M^k - M^\infty)$ is a constant). Consideration of 3 consecutive points and elimination of the constant leads to the following formula for the extrapolation of individual matrix elements:

$$(29) \quad M^{k+1} = M^{k+1} + \Delta M^{k+1} / (\Delta M^k - \Delta M^{k+1}), \quad \Delta M^k = M^k - M^{k-1}$$

This formula can be applied to the MO coefficients, density matrix or Fock matrix to improve convergence in the later stages of an SCF calculation. In addition, the Aitken extrapolation has been shown to overcome oscillatory divergence in the SCF²⁴. A variety of other extrapolation schemes have also been used²⁵⁻²⁹. As with the damping, orthonormality and idempotency are not maintained²⁰; hence the final iteration must be carried out without extrapolation to obtain a variational energy and orthonormal MO's.

Univariate Search A simple SCF cycle provides a direction in which the orbital coefficient matrix or the density matrix is to be changed, and a step length in that direction. Instead of using the simple SCF step length, better results can be obtained by minimizing the energy with respect to the step size. Care must be taken to ensure that the MO's remain orthonormal as the step size is varied. If the minimum in the univariate search is to be

located accurately³⁰⁻³³, the energy can be calculated at several points along the search path until the minimum is bracketed. A minimum is then found by fitting a polynomial to the energies along the search path. An approximate location of the minimum in the search direction can be found with much less effort^{18,19,34} (in fact, without any additional energy or Fock matrix evaluations). A cubic polynomial can be fit to the energies, E^k and E^{k+1} , and to the derivatives $dE^k/d\Delta\mathbf{C}$ and $dE^{k+1}/d\Delta\mathbf{C}$ which can be obtained from the Fock matrices \mathbf{F}^k and \mathbf{F}^{k+1} . The new matrix (MO, density or Fock) at the minimum in the cubic polynomial can then be found by interpolation. Optimizing the damping factor in eq 26 is a special case of this approach^{18,19}. Experience with general unconstrained non-linear optimization problems has pointed out the importance of the univariate search step in obtaining rapid convergence in quasi-Newton minimizations³⁵. The same appears to be true for the constrained optimization problem posed by the Roothaan-Hall equations.

DIIS The DIIS approach³⁶ is a general and efficient method for obtaining a solution to a set of equations that must be solved iteratively. Its roots are linked to the Lánczos method³⁷ for eigenvalue problems. An iterative solution of a set of equations produces a series of vectors that approach the desired result. Initially, these vectors span a smaller space than the full set of equations. The equations can be projected into this small space and inverted directly to yield an approximate solution that is used as a better input for the next iteration. Specifically, at the k^{th} iteration, a better approximation to the solution is obtained that is of the form $\mathbf{p} = \sum c_i \mathbf{p}^i$ subject to the constraint $\sum c_i = 1$ such that the error $\sum c_i \mathbf{e}^i$ is minimized in a least squares sense. This leads to the following set of equations:

$$(30) \quad \begin{bmatrix} 0 & -1 \\ -1^T & \mathbf{b} \end{bmatrix} \begin{bmatrix} -\lambda \\ \mathbf{c} \end{bmatrix} = \begin{bmatrix} -1 \\ \mathbf{0} \end{bmatrix}$$

where $\mathbf{1} = (1, 1, \dots)$, $b_{ij} = \text{tr } \mathbf{e}^{i\dagger} \mathbf{e}^j$ and \mathbf{e}^i is an estimate of the error at the i^{th} iteration. In one form or another, this approach has been used for geometry optimization³⁸, CI calculations³⁹, CPHF equations⁴⁰ and SCF iterations^{36,41}. For SCF calculations, the occupied-virtual block of the Fock operator in the MO basis could be used as the error vector^{36a}. A better choice for the error vector has been found to be^{36b}

$$(31) \quad \mathbf{e}^k = (\mathbf{S}^{-1/2})^T (\mathbf{F}^k \mathbf{D}^k - \mathbf{D}^k \mathbf{F}^k) \mathbf{S}^{-1/2}$$

The linear combination coefficients, c_i , are determined by eq 30 (if the equation becomes ill conditioned, results from the earliest iterations are discarded one at a time until the equations are well behaved again). An improved Fock matrix is constructed according to $\mathbf{F} = \sum c_i \mathbf{F}^i$, which is diagonalized to obtain the coefficients for the next iteration.

The DIIS approach can be thought of as a multivariate search and hence should be significantly better than the univariate search methods. Applications indicate that DIIS overcomes most of the convergence problems of simple SCF³⁶. Damping or level shifting can be used with DIIS to overcome most of the remaining convergence problems. The DIIS approach is often a factor of 2 or 3 faster than simple SCF³⁶ and, in a number of cases^{36b}, has been shown to be twice as fast as quadratically convergent SCF⁴².

Quadratically Convergent SCF (QC-SCF)⁴² One source of the difficulties with the simple SCF iteration scheme is the fact that the Fock matrix depends on the MO coefficients. A quadratically convergent SCF procedure remedies this problem, in part, by taking this dependency into account through first order. As discussed in the Stability section, the Hartree-Fock wavefunction is invariant with respect to the occupied-occupied and virtual-virtual orbital mixing. Hence only the effect of occupied-virtual mixing on the energy needs to be considered. As in the stability calculation, small contributions from the virtual orbitals are added to the occupied orbitals, eq 17. The wavefunction and the total energy are expressed in terms of the occupied-virtual mixing coefficients, D_{ia} , shown in eq 17 - 20. Since the mixing coefficients are no longer infinitesimal, the denominator must be retained through second order. The variational energy can then be written as:

$$(32) \quad E = E_0 + \left\{ \begin{bmatrix} \mathbf{D} \\ \mathbf{D}^* \end{bmatrix}^\dagger \begin{bmatrix} \mathbf{D}_0 \mathbf{F} \\ \mathbf{D}_0^* \mathbf{F}^* \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \mathbf{D} \\ \mathbf{D}^* \end{bmatrix}^\dagger \begin{bmatrix} \mathbf{A} & \mathbf{D}_0 \mathbf{B} \\ \mathbf{D}_0^* \mathbf{B}^* & \mathbf{A}^* \end{bmatrix} \begin{bmatrix} \mathbf{D} \\ \mathbf{D}^* \end{bmatrix} \right\} \{ \mathbf{D}_0^* \mathbf{D}_0 + \mathbf{D}^\dagger \mathbf{D} \}^{-1}$$

where $\mathbf{F} = \langle \Psi_1^a | \hat{H} | \Psi_0 \rangle$, the occupied-virtual block of the Fock matrix. Since $\mathbf{D}_0 \approx 1$ near convergence, this expression can be simplified by dropping the \mathbf{D}_0 dependence of the terms in \mathbf{B} to obtain an equation that is quadratic in the \mathbf{D} 's. The energy can now be written as

$$(33) \quad E = E_0 + \{ \tilde{\mathbf{D}}^\dagger \tilde{\mathbf{F}} \tilde{\mathbf{D}}_0 + \frac{1}{2} \tilde{\mathbf{D}}^\dagger \mathbf{H} \tilde{\mathbf{D}} \} \{ \frac{1}{2} \tilde{\mathbf{D}}_0^\dagger \tilde{\mathbf{D}}_0 + \frac{1}{2} \tilde{\mathbf{D}}^\dagger \tilde{\mathbf{D}} \}^{-1}$$

$$\text{where } \tilde{\mathbf{D}}_0 = \begin{bmatrix} \mathbf{D}_0 \\ \mathbf{D}_0^* \end{bmatrix} \text{ and } \tilde{\mathbf{F}} = \begin{bmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^* \end{bmatrix}$$

The energy must be stationary with respect to variations in the \mathbf{D} 's. Hence

$$(34a) \quad \frac{dE}{d\tilde{\mathbf{D}}_0} = \{ \tilde{\mathbf{D}}^\dagger \tilde{\mathbf{F}} - (E - E_0) \tilde{\mathbf{D}}_0^\dagger \} \{ \frac{1}{2} \tilde{\mathbf{D}}_0^\dagger \tilde{\mathbf{D}}_0 + \frac{1}{2} \tilde{\mathbf{D}}^\dagger \tilde{\mathbf{D}} \}^{-1} = 0$$

$$(34b) \quad \frac{dE}{d\tilde{\mathbf{D}}^\dagger} = \{ \tilde{\mathbf{F}} \tilde{\mathbf{D}}_0 + \mathbf{H} \tilde{\mathbf{D}} - (E - E_0) \tilde{\mathbf{D}} \} \{ \frac{1}{2} \tilde{\mathbf{D}}_0^\dagger \tilde{\mathbf{D}}_0 + \frac{1}{2} \tilde{\mathbf{D}}^\dagger \tilde{\mathbf{D}} \}^{-1} = 0$$

This can be written conveniently in matrix form as an eigenvalue problem:

$$(35) \quad \begin{bmatrix} E_0 \mathbf{I} & \tilde{\mathbf{F}}^\dagger \\ \tilde{\mathbf{F}} & E_0 \mathbf{I} + \mathbf{H} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{D}}_0 \\ \tilde{\mathbf{D}} \end{bmatrix} = E \begin{bmatrix} \tilde{\mathbf{D}}_0 \\ \tilde{\mathbf{D}} \end{bmatrix}$$

with the eigenvector normalized such that $\{ \frac{1}{2} \tilde{\mathbf{D}}_0^\dagger \tilde{\mathbf{D}}_0 + \frac{1}{2} \tilde{\mathbf{D}}^\dagger \tilde{\mathbf{D}} \} = 1$. In the case of a wavefunction constrained to be real, eq 35 reduces to the equation presented by Bacskay⁴².

$$(36) \quad \begin{bmatrix} E_0 & \mathbf{F}^\dagger \\ \mathbf{F} & E_0 \mathbf{I} + \mathbf{A} + \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{D}_0 \\ \mathbf{D} \end{bmatrix} = E \begin{bmatrix} \mathbf{D}_0 \\ \mathbf{D} \end{bmatrix}; \quad \mathbf{D}_0^2 + \mathbf{D}^\dagger \mathbf{D} = 1$$

If \mathbf{B} and the two electron part of \mathbf{A} are omitted, the QC-SCF method is the same as the simple SCF⁴².

At first glance, eq 35 and 36 would seem to be very expensive to solve, since each iteration involves a 4 index transformation of the 2 electron integrals and requires a large matrix to be diagonalized to obtain the \mathbf{D} 's. However, Bacskay⁴² has shown that the

eigenvalue problem can be solved iteratively and efficiently by avoiding the explicit 4 index transformation. The products $\mathbf{H}\bar{\mathbf{D}}$ or \mathbf{AD} and \mathbf{BD} can be constructed from the untransformed integrals⁴². Thus one micro-iteration of the solution of the eigenvalue problem requires about the same amount of work as one iteration of the conventional SCF. Far fewer macro-iterations are needed to converge the QC-SCF, with the net result that convergence is reached in significantly less cpu time than the simple SCF. Although DIIS is generally faster^{36b} than QC-SCF, the latter may have some advantages for difficult to converge cases or where very tight convergence is needed. An alternative approach to QC-SCF has been presented by Douady, Elinger, Subra and Levy⁴³ based on exponential transformation of the MO coefficients.

Relation Between Stability and Convergence Whether or not a simple SCF calculation converges is closely related to the stability conditions⁴⁴⁻⁴⁶. An SCF calculation will converge if the changes in the MO's decrease from one cycle to the next. Let the changes be written as

$$(37) \quad \Delta \mathbf{M}^{k+1} = \mathbf{Q} \Delta \mathbf{M}^k$$

The iterative process will converge only if the absolute values of the eigenvalues of \mathbf{Q} are less than unity. As before (eq 17 or 27), it is sufficient to consider only the occupied-virtual orbital mixing in determining the changes in the orbitals. Let $\mathbf{C}_{\mu p}^k$ be the eigenvectors of \mathbf{F}^k . These eigenvectors can be used to construct \mathbf{F}^{k+1} according to eq 24. The coefficients and density matrix can be related to the corresponding quantities from the previous iteration:

$$(38) \quad \mathbf{C}_{vi}^k = \mathbf{C}_{vi}^{k-1} + \sum_a \mathbf{D}_{ia}^k \mathbf{C}_{va}^k$$

$$(39) \quad \mathbf{P}_{\lambda\sigma}^k = \mathbf{P}_{\lambda\sigma}^{k-1} + \sum_{ia} [\mathbf{D}_{ia}^{k*} \mathbf{C}_{\lambda a}^k \mathbf{C}_{\sigma i}^k + \mathbf{D}_{ia}^k \mathbf{C}_{\lambda i}^{k*} \mathbf{C}_{\sigma a}^k]$$

where only the first order terms in \mathbf{D} have been retained. Since \mathbf{F}^k was built from \mathbf{P}^{k-1} , only the contribution from $\mathbf{P}^k - \mathbf{P}^{k-1}$ needs to be added to \mathbf{F}^k to obtain \mathbf{F}^{k+1} . Inserting eq 39 into eq 24 and transforming to the $\mathbf{C}_{\mu p}^k$ basis yields

$$(40) \quad \mathbf{F}_{pq}^{k+1} = \mathbf{F}_{pq}^k + \sum_{jb} [(\text{pjllqb}) \mathbf{D}_{jb}^k + (\text{pblqj}) \mathbf{D}_{jb}^{k*}] \\ = \varepsilon_p \delta_{pq} + \sum_{jb} [(\text{pjllqb}) \mathbf{D}_{jb}^k + (\text{pblqj}) \mathbf{D}_{jb}^{k*}]$$

since the $\mathbf{C}_{\mu p}^k$ are the the eigenvectors of \mathbf{F}^k . With the definitions in eq 21 and 22, the occupied-virtual block of \mathbf{F}^{k+1} can be written as

$$(41) \quad \mathbf{F}^{k+1} = (\mathbf{A} - \Delta) \mathbf{D}^k + \mathbf{B} \mathbf{D}^{k*}$$

where $\Delta_{ia,jb} = (\varepsilon_a - \varepsilon_i) \delta_{ij} \delta_{ab}$. The new coefficients, $\mathbf{C}_{\mu i}^{k+1}$, (eq 24) can be obtained by diagonalizing \mathbf{F}^{k+1} (see eq 25); alternatively the \mathbf{D} 's can be computed by perturbation theory (see eq 28):

$$(42) \quad D_{ia}^{k+1} = -F_{ai}^{k+1} / (\epsilon_a - \epsilon_i) = - \sum_{jb} [(ajlib) D_{jb}^k + (ablij) D_{jb}^{k*}] / (\epsilon_a - \epsilon_i)$$

$$\text{or} \quad \tilde{D}^{k+1} = -\tilde{\Delta}^{-1} (\tilde{H} - \tilde{\Delta}) \tilde{D}^k; \quad \tilde{\Delta} = \begin{bmatrix} \Delta & 0 \\ 0 & \Delta \end{bmatrix}$$

After a bit of rearrangement, eq 42 can be put in the form of eq 37.

$$(43) \quad \tilde{\Delta}^{-1/2} \tilde{D}^{k+1} = (1 - \tilde{\Delta}^{-1/2} \mathbf{H} \tilde{\Delta}^{-1/2}) \tilde{\Delta}^{1/2} \tilde{D}^k$$

For the iteration to converge, the absolute values of the eigenvalues of $Q = (1 - \tilde{\Delta}^{-1/2} \mathbf{H} \tilde{\Delta}^{-1/2})$ must be less than 1. Hence, the eigenvalues of $\tilde{\Delta}^{-1/2} \mathbf{H} \tilde{\Delta}^{-1/2}$ must be between 0 and 2⁴⁷. Since $\tilde{\Delta}$ is diagonal and positive definite, \mathbf{H} must at least be positive definite for the simple SCF to converge. But if \mathbf{H} is positive semidefinite, then the wavefunction is stable. Thus, if it converges, the simple SCF procedure (without techniques such as damping, level shifting, etc.) yields a stable wavefunction. However, the converse is not necessarily true. If a wavefunction is stable, \mathbf{H} is positive semidefinite; however $\tilde{\Delta}^{-1/2} \mathbf{H} \tilde{\Delta}^{-1/2}$ may have eigenvalues larger than 2 and the simple SCF does not converge. If any of the eigenvalues of $\tilde{\Delta}^{-1/2} \mathbf{H} \tilde{\Delta}^{-1/2}$ are near 0 or 2, convergence will be slow. An eigenvalue near zero signals a near degeneracy between two states. Large eigenvalues of $\tilde{\Delta}^{-1/2} \mathbf{H} \tilde{\Delta}^{-1/2}$ can occur when an energy difference between occupied and virtual orbitals is small relative to the 2 electron contributions to \mathbf{H} . This gives rise to large changes in the Fock matrix from step to step and results in oscillatory divergence if any of the eigenvalues of $\tilde{\Delta}^{-1/2} \mathbf{H} \tilde{\Delta}^{-1/2}$ are greater than 2. This analysis also illustrates how level shifting affects the convergence of a conventional SCF calculation. Since level shifting increases all of the occupied-virtual orbital energy differences (i.e. all of the elements of Δ become larger), it reduces the eigenvalues of $1 - \tilde{\Delta}^{-1/2} \mathbf{H} \tilde{\Delta}^{-1/2}$ and forces convergence, even in cases that would otherwise be divergent.

Examples The behavior of some of the algorithms discussed above is shown in Table 3. The convergence is judged by the root mean square change in the density matrix (or the rms orbital rotation gradient in the case of QCSCF). The number of Fock matrix evaluations (and number of micro-iterations to solve QCSCF equations) is used as the iteration count. Since most of the work in the SCF for a larger molecule is in the formation of the Fock matrix, this method of counting iterations gives a fair comparison of the work required to achieve convergence (a micro-iteration of the QCSCF equations requires about the same amount of work as a Fock matrix evaluation⁴²). For the simple SCF, the energy was constant to 10 significant figures after 10 iterations, but the density matrix continued to change. The SCF seemed to be heading toward convergence on the density matrix for the first 18 cycles; however, it then started to diverge. Extrapolation did not alter this behavior significantly. Level shifting slowed the rate of convergence, but did not prevent divergence. A univariate search with accurate location of the minimum, the SCFDM approach detailed in reference 32, did lead to convergence after 9 searches which required a total of 65 Fock matrix evaluations. By contrast, the DIIS algorithm converged very rapidly, in only 8 cycles. The QCSCF method started with a univariate search that required 7 points to find the minimum. This was followed by 2 QCSCF macro-iterations, each

Table 3. Convergence behavior of various SCF algorithms for O₂^a

Iteration	Simple SCF ^{b,c}	Level Shift ^{b,d}	SCFDM ^e	DIIS	QCSCF ^f
1					6.7(-1)
2	4.9(-4)	4.6(-4)		3.0(-4)	
3	1.7(-4)	1.8(-4)		1.8(-4)	
4	7.4(-5)	8.2(-5)		1.2(-4)	
5	3.5(-5)	4.1(-5)	4.1(-3)	2.0(-5)	
6	1.7(-5)	2.2(-5)		9.6(-9)	
7	8.5(-6)	1.1(-5)			
8	4.2(-6)	6.0(-6)	6.6(-4)	1.9(-11)	4.5(-3)
9	2.1(-6)	3.2(-6)			
10	1.0(-6)	1.7(-6)			
11	5.1(-7)	9.0(-7)	5.3(-4)		1.6(-6)
12	2.5(-7)	4.8(-7)			
13	1.3(-7)	2.5(-7)			
14	6.2(-8)	1.3(-7)			2.5(-10)
15	3.1(-8)	7.1(-8)			
16	1.5(-8)	3.8(-8)			
17	8.5(-9)	2.0(-8)			
18	8.2(-9)	1.1(-8)	3.9(-4)		
19	1.4(-8)	5.9(-9)			
31	3.5(-5)	1.4(-6)	5.5(-6)		

^a UHF triplet at R = 1.7Å; convergence on root mean square change in the density matrices; the iteration count indicates the number of Fock matrix evaluations (and micro-iterations required to solve the QCSCF equations)

^b failed to converge in 32 cycles.

^c 3 and 4 point extrapolation did not alter convergence significantly.

^d virtual orbital energies shifted by 0.1 au.

^e ref 32, univariate search with accurate location of the minimum; converged after 65 Fock matrix evaluations.

^f values quoted in this column are rms orbital rotation gradients rather than rms density matrix changes.

requiring 3 micro-iterations to solve the QCSCF equations, and a final energy and rotation gradient calculation, for a total equivalent to 14 Fock matrix evaluations. An accurate univariate search, both in QCSCF and SCFDM, imposes a large penalty in terms of the total number of Fock matrix evaluations. This suggests that an approximate search might be more cost effective; however, this may risk non-convergence. The DIIS method is clearly the best algorithm for this example; similar results have been demonstrated for other cases³⁶. QCSCF also performs quite well and should be used if DIIS fails. Simple SCF is not satisfactory for this example, even with various convergence controlling techniques.

5. Initial Guess.

Whether or not an SCF calculation converges and how many iterations it takes to converge often depends on the quality of the initial guess for the MO's or the density matrix. Some of the approaches used to obtain a suitable initial guess are listed below.

- (a) Eigenvectors of the one electron hamiltonian ($h_{\mu\nu}$ in eq 9). This is equivalent to choosing a zero density matrix in building the Fock matrix for the zeroth iteration.
- (b) Semi-empirical wavefunction. Extended Hückel (EHT), CNDO, MNDO etc. provide reasonable MO's; the empirical parameters in the EHT are sometimes adjusted until more suitable orbitals are obtained.
- (c) Smaller basis set ab initio calculations. These can be projected onto larger basis sets to yield good initial guesses.
- (d) Calculations from near by points. This is useful for geometry optimizations, numerical differentiations and potential surface scans.

The above methods are roughly in order of increasing quality. For problematic cases, a number of additional techniques and tricks can be taken to obtain convergence:

- (e) Alter the orbital occupancy. Selected occupied and virtual orbitals can be switched until a combination is found that leads to convergence for the desired state. Symmetry and a close inspection of the orbitals usually indicates which orbitals need to be switched.
- (f) Higher symmetry. Often, higher symmetry structures are better behaved, since symmetry reduces the orbital mixing that occurs. The converged orbitals can serve as an initial guess for near by lower symmetry structures.
- (g) Localized orbitals. It may be easier to choose the correct occupancy if the orbitals are not delocalized over the entire molecule.
- (h) Changing the geometry. Elongated bonds, strained structures and strongly interacting fragments sometimes cause stability and convergence problems. It may be possible to converge the SCF calculation for a structure with shorter bonds, a more relaxed geometry or non-interacting fragments, respectively (however, care should be taken not to change the symmetry of the system or its alignment with the coordinate axes). These orbitals can then be used as a guess for the problematic case.
- (i) Stability analysis. If a stability calculation indicates that a converged wavefunction is unstable, the wavefunction can be displaced along the offending eigenvector and optimized by a univariate search, to give an initial guess for a lower energy wavefunction.
- (j) Changing the charge. It may be possible to converge a calculation with fewer electrons (but the same number of unpaired electrons) or with greater charges on the nuclei. The electrons in a cation are more strongly bound and hence more likely to converge. The converged orbitals may be suitable as an initial guess for the unaltered molecule.

There are, of course, cases for which the best initial guess will not lead to SCF convergence with a given SCF procedure. These can only be overcome by using a better algorithm. With a combination of improvements in the initial guess and the use of more powerful algorithms, most SCF convergence problems can be overcome.

6. Summary.

Stability SCF calculations don't always converge to the most stable wavefunction. Molecules with unusual geometric or electronic structure (e.g. strained geometry, transition states, hypervalency, open shells etc.) can have internal and/or external instabilities in the wavefunction, and therefore should be tested for stability. If necessary, a lower energy wavefunction can be found by changing the MO's according to the offending eigenvector obtained in the stability analysis.

Convergence Getting a converged SCF wavefunction can sometimes be a problem. The methods discussed above can be summarized as follows:

- (a) If at all possible, avoid the simple SCF iteration scheme. DIIS is more stable, requires fewer iterations and costs no more per iteration than simple SCF.
- (b) Use a good initial guess - either from a reliable semi-empirical calculation, a lower level ab initio calculation or a near by structure. Sometimes occupied and virtual orbitals have to be switched or mixed to get a suitable initial guess.
- (c) Damping may be useful in getting a partially converged wavefunction, but level shifting is usually more efficient. Univariate search is even better but may be more expensive.
- (d) Extrapolation can speed up convergence during the latter part of a simple SCF calculation, but may cause problems if applied too early (DIIS normally does not need additional extrapolation).
- (e) Quadratically convergent SCF is usually more expensive than DIIS, but may be the method of choice if DIIS fails; however, the initial guess must be suitably close to the desired wavefunction (see (c) to get close enough).

References.

1. A. Szabo and N.S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Macmillan, New York, 1982.
2. R. McWeeny and B.T. Sutcliffe, *Methods of Molecular Quantum Mechanics*, Academic Press, London, 1969.
3. J.P Lowe, *Quantum Chemistry*, Academic Press, New York, 1978.
4. C. C. J. Roothaan, *Rev. Mod. Phys.*, **23**, 69 (1951); G. G. Hall, *Proc. Roy. Soc. (London)*, **A205**, 541 (1951)
5. J. A. Pople and R. K. Nesbet, *J. Chem. Phys.* **22**, 571 (1954).
6. D. J. Thouless, *The Quantum Mechanics of Many-Body Systems*, Academic, NY, 1961.
7. J. Cizek and J. Paldus, *J. Chem. Phys.* **47**, 3976 (1967), *Phys. Rev A*, **3**, 525 (1971); J. Paldus and J. Cizek, *Chem. Phys. Lett.*, **3**, 1 (1969), *J. Chem. Phys.*, **52**, 2919 (1970).
8. N. S. Ostlund, *J. Chem. Phys.* **57**, 2994 (1972).
9. H. Fukutome, *Int. J. Quantum Chem.* **20**, 955 (1981).
10. R. Seeger and J.A. Pople, *J. Chem. Phys.*, **66**, 3045, (1977).
11. UHF wavefunctions can be projected⁵⁶ to make them eigenfunctions of \hat{S}^2 but the result is multideterminantal and, hence, outside the scope of this article.
12. P. -O. Löwdin, *Phys. Rev.*, **97**, 1509 (1955); for a more recent review see I. Mayer, *Adv. Quantum Chem.*, **12**, 189 (1980).

13. For ROHF wavefunctions, singly and doubly occupied orbitals must also be mixed.
14. The stabilities of projected Hartree-Fock methods are discussed in reference 15
15. Y. G. Smeyers and G. Delgado-Barrio, *Phys. Rev. A*, **15**, 1359 (1977); P. Karadakov and O. Castano, *Int. J. Quantum Chem.*, **22**, 1033 (1982); O. Castano and P. Karadakov, *Chem. Phys. Lett.*, **130**, 123 (1986).
16. D. R. Hartree, *The Calculation of Atomic Structures*, Wiley, New York, 1957; C. C. J. Roothaan and P. S. Bagus, in *Methods in Computational Physics*, Vol 2, B. Adler, S. Fernbach and M. Rotenberg, eds., Academic Press, New York, 1963.
17. M.C. Zerner and M. Hehenberger, *Chem. Phys. Lett.*, **62**, 550, (1979).
18. G. Karlström, *Chem. Phys. Lett.*, **67**, 348, (1979).
19. P. de Montgolfier and A. Hoareau, *J. Chem. Phys.* **65**, 2477 (1976).
20. For methods of maintaining idempotency while changing the density matrix see ref. 30.
21. V. R. Saunders and I. H. Hillier, *Int. J. Quantum Chem.* **7**, 699 (1973); M. F. Guest and V. R. Saunders, *Mol. Phys.*, **28**, 819 (1974).
22. A. V. Mitin, *J. Comput. Chem.* **9**, 107 (1988).
23. For a discussion of Aitken extrapolation and related methods see W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes. The Art of Scientific Computing*, Cambridge Univ. Press, Cambridge, 1986, or any similar volume on numerical analysis.
24. J. Cioslowski, *J. Chem. Phys.*, **89**, 2126 (1988).
25. W. B. Neilsen, *Chem. Phys. Lett.*, **18**, 225 (1973).
26. I. Röeggen, *Chem. Phys. Lett.*, **22**, 140 (1973).
27. M. J. S. Dewar and P. K. Weiner, *Comput. Chem.* **2**, 31 (1978).
28. L. G. Ferreira, *J. Comput. Phys.*, **36**, 198 (1980).
29. Z. Chen, *Chem. Phys. Lett.*, **87**, 455 (1982).
30. R. McWeeny, *Proc. Roy. Soc. (London)* **A235**, 496 (1956).
31. I.H. Hillier and V.R. Saunders, *Int. J. Quantum Chem.*, **4**, 503, (1970), *Proc. Roy. Soc. (London)*, **A320**, 161 (1970).
32. R. Seeger and J. A. Pople, *J. Chem. Phys.* **65**, 265 (1976).
33. R. N. Camp and H. F. King, *J. Chem. Phys.* **75**, 268 (1981).
34. J. F. Rico, J. M. G. de la Vega, M. Paniagua and J. I. Fernández-Alonso, *J. Chem. Phys.* **79**, 4407 (1983).
35. L. E. Scales, *Introduction to Non-Linear Optimization*, Springer, New York, 1985.
36. P. Pulay, *Chem. Phys. Lett.* **73**, 393 (1980); *J. Comput. Chem.* **3**, 556 (1982).
37. C. Lanczos, *J. Res. Natl. Bur. Std. US*, **45**, 255 (1955).
38. P. Császár and P. Pulay, *J. Mol. Struct.*, **114**, 31 (1984).
39. E. R. Davidson, *J. Comput. Phys.* **17**, 87 (1975).
40. J. A. Pople, R. Krishnan, H. B. Schlegel and J. S. Binkley, *Int. J. Quantum Chem., Quantum Chem. Symp.*, **13**, 225 (1979).
41. I. Bálint and M. I. Bán, *Chem. Phys. Lett.* **101**, 153 (1983).
42. G. B. Bacskay, *Chem. Phys.* **61**, 385 (1981), **65**, 383 (1982).
43. J. Douady, Y. Elinger, R. Subra and B. Levy, *J. Chem. Phys.* **72**, 1452 (1980).
44. R. E. Stanton, *J. Chem. Phys.* **75**, 3426, 5416 (1981).
45. J. C. Facelli and R. H. Contreras, *J. Chem. Phys.* **77**, 1076 (1982); **79**, 3421 (1983).
46. M. A. Natiello and G. E. Scuseria, *Int. J. Quantum Chem.*, **26**, 1039 (1984).
47. with the appropriate change in notation, this result is the same as reference 45.