The Self-Consistent Field Equations for Generalized Valence Bond and Open-Shell Hartree–Fock Wave Functions

Frank W. Bobrowicz and William A. Goddard III

1. Introduction

The basic starting point for calculating \textit{ab initio} wave functions of molecules is generally the Hartree–Fock (HF) wave function, which in the simplest case involves two electrons (one with each spin) in each orbital $\phi_i$ with the total wave function antisymmetrized in order to satisfy the Pauli principle

$$a[(\phi_1 \alpha)(\phi_1 \beta)(\phi_2 \alpha)(\phi_2 \beta) \cdots (\phi_n \alpha)(\phi_n \beta)] = a(\phi_1 \phi_2 \phi_2 \cdots \phi_n \alpha \beta \alpha \beta \cdots \alpha \beta)$$

(1)

Here $a$ is the antisymmetrizer or determinant operator* and $\alpha$ and $\beta$ are the usual spin functions. In Eq. (1) as elsewhere, we arrange products of spatial functions and spin functions in order of increasing electron numbers.

* $a = \sum \zeta_\tau$ where the sum is over all permutations $\tau$ and where $\zeta_\tau = \pm 1$ is the parity of $\tau$.
As is well known, this type of wave function generally does not lead to an adequate description of bond dissociation, nor does it provide a consistent level of treatment for excited states. A simple generalization of Eq. (1) is used to allow selected electron pairs to be described in terms of overlapping singly occupied orbitals,

$$\phi_i(1)\phi_i(2) \rightarrow [\phi_{ia}(1)\phi_{ib}(2) + \phi_{ib}(1)\phi_{ia}(2)]$$

(2)
as in the simple valence bond wave function. If the various orbitals of such a wave function are solved for self consistency, the resulting wave function is referred to as a generalized valence bond (GVB) wave function.\(^{(1-3)}\) In this chapter we will develop the variational equations which are solved to obtain GVB wave functions, and we will describe the various steps for solving the GVB wave functions of general systems. This discussion necessarily includes open-shell Hartree–Fock wave functions as a special case and the procedures developed are also appropriate for more general wave functions.

A common procedure for solving SCF wave functions\(^{(4)}\) has been to consider all occupied orbitals as fixed except for one, say $\phi_k$, and to require that the first-order change in the energy be zero upon variation of this orbital, consistent with maintaining orthogonality between all occupied orbitals. A typical approach is to write the variational equations as pseudoeigenvalue equations and to solve these equations iteratively. Although often successful, this procedure is not reliable for complicated wave functions.

Our approach is to start with the expression for the total energy and to expand it (through second order) in terms of simultaneous corrections in all orbitals.\(^{(5,6)}\) Because of the orthogonality conditions the variations in different orbitals are coupled. Including these couplings, we derive the equations for the optimal simultaneous correction in each orbital. Solution of these equations leads to quadratic convergence to the optimum orbitals. Various simplifications are also discussed.

2. The Energy Expression

Given the usual nonrelativistic Hamiltonian

$$H = \sum_{p=1}^{N} h(p) + \sum_{p>q=1}^{N} \frac{1}{r_{pq}}$$

(3)

where $N$ is the number of electrons and the operator $h(p)$ contains all one-electron terms involving electron $p$, the total energy of any electronic wave function can be written as\(^{(7)}\)

$$E = \sum_{i,j} D_{ij}^j h_{ij} + \sum_{i,j,k,l} D_{ijkl}^k (i) (k | j | l)$$

(4)
where

\[ h_{ij} = \langle \phi_i | h | \phi_j \rangle \]
\[ (ik|jl) = \langle \phi_i (1) \phi_j (2) | (1/r_{12}) | \phi_k (1) \phi_l (2) \rangle \]

and \( \{ \phi_i \} \) is a set of \( n \) orbital basis functions in terms of which the wave function is described. Here \( D_i^j \) and \( D_k^l \) are appropriate density matrix elements.

For a wide class of useful wave functions (including open-shell HF and important cases of GVB), once appropriately expressed in terms of orthonormal orbitals

\[ \langle \phi_i | \phi_j \rangle = \delta_{ij} \] (5)

the one- and two-electron density matrices in (4) can be brought into diagonal form so that the only nonzero elements are

\[ D_i^j = 2f_i \quad D_i^j = a_{ij} \quad D_k^l = b_{kl} \]

where the energy coefficients \( \{ f_i, a_{ij}, b_{kl} \} \) are independent of the orbitals \( \{ \phi_i \} \). For such wave functions the energy expression takes on the greatly simplified form,

\[ E = 2 \sum_{i=1}^{n} f_i h_{ii} + \sum_{i,j=1}^{n} (a_{ij} J_{ij} + b_{ij} K_{ij}) \] (6)

where

\[ J_{ij} = (ii|jj) = \langle \phi_i | J | \phi_i \rangle = \langle \phi_j | J_i | \phi_j \rangle \]
\[ K_{ij} = (ij|ij) = \langle \phi_i | K_i | \phi_j \rangle = \langle \phi_j | K_i | \phi_j \rangle \]

indicate the usual Coulomb and exchange energies, respectively. The Coulomb and exchange operators \( J_i \) and \( K_i \) are defined here for later convenience. In this chapter we will be concerned mainly with the problem of solving for the self-consistent solutions of wave functions whose energy expressions conform to Eq. (6).

3. Wave Functions

While Eq. (6) does not apply to all electronic wave functions, this energy expression is applicable to many of the most important wave functions useful in dealing with chemical problems at the conceptual level. In this section we will examine such wave functions, paying particular attention to both the simplifying aspects of their functional forms and their inherent limitations.
3.1. Hartree–Fock Wave Functions

The HF wave function is of singular importance in our basic understanding of atomic and molecular electronic structure. For mainly historical reasons, however, this wave function defies a neat unambiguous definition, since a fairly wide variety of wave functions of differing forms are classified under this appellation. Therefore, we will discuss HF wave functions by way of representative examples, without attempting an exhaustive treatment.

3.1.1. Closed Shell

Consider the four-electron singlet wave function involving two doubly occupied orbitals,

\[ a[\phi_1, \alpha \phi_1 \beta \phi_2 \alpha \phi_2 \beta] = a[\phi_1, \phi_1 \phi_2 \phi_2 \alpha \beta \alpha \beta] \]  

(7)

Allowing orbitals \( \phi_1 \) and \( \phi_2 \) to be nonorthogonal,

\[ \langle \phi_1 | \phi_2 \rangle = S_{12}, \quad \langle \phi_1 | \phi_1 \rangle = \langle \phi_2 | \phi_2 \rangle = 1 \]  

(8)

the energy of (7) is

\[ E = [2(1 - S_{12}^2)(h_{11} + h_{22} - 2S_{12}^2h_{12}) + J_{11} + J_{22} + 2(2 - S_{12}^2)J_{12} - 2(1 - 3S_{12}^2)K_{12} \]

\[ - 4S_{12} \langle \phi_1 | (J_1 + J_2) | \phi_2 \rangle ]/(1 - S_{12}^2)^2 \]  

(9)

However, writing \( \phi_2 \) as

\[ \phi_2 = \phi'_2 + \lambda \phi_1 \]

where

\[ \langle \phi_1 | \phi'_2 \rangle = 0 \]  

(10)

(7) becomes

\[ a[\phi_1, \phi_1 \phi_2 \phi_2 \alpha \beta \alpha \beta] = a[\phi_1, \phi_1 \phi'_2 \phi'_2 \alpha \beta \alpha \beta] + \lambda a[\phi_1, \phi_1 \phi'_2 \phi_2 \alpha \beta \alpha \beta] \]

\[ + \lambda^2 a[\phi_1, \phi_1 \phi_2 \phi_2 \alpha \beta \alpha \beta] \]  

(11)

But, because of the Pauli principle (as embodied in the antisymmetrizer \( a \)), we find that

\[ a[\cdots \phi_1 \alpha \cdots \phi_1 \alpha \cdots ] = a[\cdots \phi_2 \beta \cdots \phi_2 \beta \cdots ] = 0 \]  

(12)

Therefore, (11) reduces to simply

\[ a[\phi_1, \phi_1 \phi_2 \phi_2 \alpha \beta \alpha \beta] = a[\phi_1, \phi_1 \phi'_2 \phi'_2 \alpha \beta \alpha \beta] \]  

(13)
That is, the antisymmetrizer projects away any nonorthogonalities, and hence it is no restriction to take these orbitals as orthogonal in the first place.* In this case the energy expression (9) becomes

$$E_c = 2h_{11} + 2h_{22} + J_{11} + J_{22} + 4J_{12} - 2K_{12}$$

(14)

Since

$$J_{ii} = K_{ii}$$

(15)

we can add and subtract these self-terms to obtain the more symmetric expression

$$E_c = 2h_{11} + 2h_{22} + (2J_{11} - K_{11}) + (2J_{22} - K_{22}) + (4J_{12} - 2K_{12})$$

(16)

which is of the form of (6) with

$$f_1 = f_2 = 1$$

$$a_{11} = a_{22} = a_{12} = 2$$

$$b_{11} = b_{22} = b_{12} = -1$$

Since $\phi_1$ and $\phi_2$ are maximally occupied ($f_i = 1$), we refer to them as closed-shell orbitals. Orbitals leading to $f_i \neq 1$ are referred to collectively as open-shell orbitals. A closed-shell orbital can be (and always is) taken to be orthogonal to all other orbitals, without restriction. A wave function composed entirely of closed-shell orbitals

$$A[\phi_1 \alpha \phi_1 \beta \phi_2 \alpha \phi_2 \beta \cdots \phi_n \alpha \phi_n \beta]$$

(17)

is called a closed-shell wave function and has the general energy expression†

$$E_c = \sum_i^2 h_{ii} + \sum_{i,j} (2J_{ij} - K_{ij})$$

(18)

3.1.2. Closed-Shell/Open-Shell Energy Partitioning

Generally, wave functions involving open-shell orbitals also contain several closed-shell orbitals as well. Therefore, it is convenient to develop a

*Also because of the antisymmetrizer we find that starting with two orbitals $\phi_1$ and $\phi_2$ and transforming to a new set of orbitals $\tilde{\phi}_1$ and $\tilde{\phi}_2$,

$$\tilde{\phi}_i = \sum_i \phi_i U_{ii}$$

leads to

$$A[\tilde{\phi}_1 \tilde{\phi}_1 \tilde{\phi}_2 \tilde{\phi}_2 \alpha \beta \alpha \beta] = (\det U)^2 A[\phi_1 \phi_1 \phi_2 \phi_2 \alpha \beta \alpha \beta]$$

Thus the closed-shell wave function is invariant under orthogonal transformations.

†We have seen that the form of (18) is invariant under orthogonal transformations of the orbitals.
scheme whereby energy expressions may be derived without explicitly considering the closed-shell orbitals. Consider the doublet wave function

\[ a[\phi_1 \phi_1 \phi_2 \phi_2 \alpha \beta \beta \alpha] \] (19)

Since all the orbitals can be taken as orthonormal, the energy of (19) is

\[ E = (2h_{11} + 2h_{22}) + h_{33} + (2J_{11} - K_{11} + 2J_{22} - K_{22} + 4J_{12} - 2K_{12}) + (2J_{13} - K_{13} + 2J_{23} - K_{23}) \] (20)

Hence

\[ f_1 = f_2 = 2f_3 = 1 \]

\[ a_{11} = a_{22} = a_{12} = 2a_{13} = 2a_{23} = 2 \]

\[ b_{11} = b_{22} = b_{12} = 2b_{13} = 2b_{23} = -1 \]

\[ a_{33} = b_{33} = 0 \]

Comparing with (16) we see that (20) can be written as

\[ E = E_c + h_{33} + 2J_{13} - K_{13} + 2J_{23} - K_{23} = E_c + h_{33} \] (21)

where

\[ h^c = h + 2J_1 - K_1 + 2J_2 - K_2 \]

Thus, insofar as the open-shell orbital (\( \phi_3 \)) is concerned all effects due to the closed-shell orbitals are included in the new "one-electron" operator \( h^c \).

In general, if orbital \( i \) is a closed-shell orbital and orbital \( j \) is any other orbital, the only nonvanishing density matrix elements in (4) involving both of these orbitals are

\[ D_{ij} = 2D_{ji}, \quad D_{ij}^* = -D_{ji}^* \]

As a consequence, the energy for any wave function involving both closed-shell and open-shell orbitals has the form

\[ E = E_c + \sum_{i,j}^{\text{open}} D_{ij}^c h_{ij}^c + \sum_{i,j}^{\text{open}} D_{ji}^c (ik|jl) \] (22)

where

\[ h^c = h + \sum_{p}^{\text{closed}} (2J_p - K_p) \]

\[ E_c = 2 \sum_{p}^{\text{closed}} h_{pp} + \sum_{p,q}^{\text{closed}} (2J_{pq} - K_{pq}) \] (23)
If the open-shell energy can be simplified to the form of (6), then (22) reduces to
\[ E = E_c + \sum_{i}^{\text{open}} 2f_i h_{ii}^c + \sum_{i j}^{\text{open}} (a_{ij} J_{ij} + b_{ij} K_{ij}) \] (24)

Hence, letting \{core\} denote the presence of an arbitrary number of closed-shell orbitals,
\[ \{\text{core}\} = \phi_{c_1} \alpha \phi_{c_2} \beta \phi_{c_3} \alpha \phi_{c_4} \beta \cdots \phi_{c_m} \alpha \phi_{c_m} \beta \] (25)

the energy for the doublet wave function
\[ a[\{\text{core}\} \phi_n \alpha] \] (26)
becomes simply
\[ E = E_c + h_{nn}^c \] (27)

3.1.3. High-Spin Open Shell

For the wave function
\[ a[\{\text{core}\} \phi_1 \phi_2 \alpha \alpha] \] (28)
where the electrons in orbitals \( \phi_1 \) and \( \phi_2 \) are coupled into a triplet state, we see from the analysis in (10)–(13) that it is no restriction to take these two orbitals as orthogonal. Therefore, the energy of (28) is
\[ E = E_c + h_{11}^c + h_{22}^c + J_{12} - K_{12} \] (29)

Adding and subtracting self-terms leads to the more symmetric expression
\[ E = E_c + h_{11}^c + h_{22}^c + \frac{1}{2}(J_{11} - K_{11}) + \frac{1}{2}(J_{22} - K_{22}) + J_{12} - K_{12} \] (30)
so that
\[ f_1 = f_2 = a_{11} = a_{22} = a_{12} = b_{11} = -b_{12} = \frac{1}{2} \]

For the general high-spin open-shell wave function,
\[ a[\{\text{core}\} \phi_1 \alpha \phi_2 \alpha \cdots \phi_n \alpha] \] (31)
we obtain the energy expression*
\[ E = E_c + \sum_i h_{ii}^c + \frac{1}{2} \sum_{i j} (J_{ij} - K_{ij}) \] (32)

We refer to wave functions such as (31) as multiplet wave functions and to the high-spin coupled open-shell orbitals involved as the multiplet-shell orbitals.

*Under a transformation \( U \) of the open-shell orbitals of a high-spin open-shell wave function, the total wave function changes as \( (\det U) \). Thus at most a sign change occurs under orthogonal transformations. As a result, the form of (32) is invariant under orthogonal transformations among the high-spin orbitals.
3.1.4. Orthogonality Restrictions

If in the open-shell singlet wave function,

$$a[(\text{core})\phi_1\phi_2(\alpha\beta - \beta\alpha)] = a[(\text{core})(\phi_1\phi_2 + \phi_2\phi_1)\alpha\beta] \quad (33a)$$

(where the electrons in $\phi_1$ and $\phi_2$ are coupled into a singlet pair), we take $\phi_1$ and $\phi_2$ as orthogonal,

$$\langle \phi_1 | \phi_2 \rangle = 0 \quad (33b)$$

the energy expression is

$$E = E_c + h_{11}^c + h_{22}^c + J_{12} + K_{12} \quad (34)$$

so that

$$f_1 = f_2 = a_{12} = b_{12} = \frac{1}{2}$$

$$a_{11} = a_{22} = b_{11} = b_{22} = 0$$

In this case, the energy expression cannot be put into a more symmetric form by adding and subtracting self-terms.

For all the HF wave functions considered in Sections 3.1.1–3.1.3, we found that because of the functional form of these wave functions the antisymmetrizer projected away any orbital nonorthogonalties. Thus, the sole effect of constraining the orbitals to be orthogonal was to permit us to express energies in the simplified form of (6), rather than (4). However, if in (33a) we allow $\phi_1$ and $\phi_2$ to be nonorthogonal,

$$\langle \phi_1 | \phi_2 \rangle = S_{12}$$

and rewrite $\phi_2$ as

$$\phi_2 = \phi_2' + \lambda \phi_1$$

where

$$\langle \phi_1 | \phi_2' \rangle = 0$$

then expansion of (33a) leads to

$$a[(\text{core})\phi_1\phi_2'(\alpha\beta - \beta\alpha)] + 2\lambda a[(\text{core})\phi_1\phi_1\alpha\beta] \quad (35)$$

The second term in (35) is nonvanishing, and hence in this case imposing orthogonality between $\phi_1$ and $\phi_2$ does, in general, represent a restriction. Thus, in this instance (33b) is not just a convenient computational tool.*

The restrictive nature of the orbital orthogonality constraint, which is always imposed upon HF wave functions, is not confined solely to the two

*The exception would be the case where orbitals $\phi_1$ and $\phi_2$ necessarily belong to different symmetry classes and hence lead to (33b) without restriction.
orbitals comprising an open-shell singlet pair. It applies in general with respect to all open-shell orbitals which are not multiplet (high-spin) coupled to one another. For example, in the wave function

$$A[\phi_1 \phi_2 \phi_3 \phi_4 \phi_5 \alpha \beta (\alpha \beta - \beta \alpha) \alpha \alpha]$$

(36)

while the conditions,

$$\langle \phi_1 | \phi_i \rangle = 0, \quad i = 2, 3, 4, 5$$
$$\langle \phi_4 | \phi_5 \rangle = 0$$

(37)

are completely unrestricted, the constraints,

$$\langle \phi_k | \phi_i \rangle = 0, \quad k = 2, 3 \text{ and } l = 4, 5$$
$$\langle \phi_2 | \phi_3 \rangle = 0$$

(38)

are actual restrictions. Whether or not (38) is imposed upon (36) can have a significant effect upon the energy (and potential usefulness) of this wave function.

### 3.2. GVB Wave Functions

#### 3.2.1. General Considerations

The Hartree–Fock wave function forms the conceptual basis for understanding the electronic structures of broad classes of molecules. However, upon applying HF wave functions to studies of chemical reactions, a serious deficiency becomes apparent. These wave functions are generally incapable of describing bond dissociation. To illustrate, consider the case of the hydrogen molecule. Near the equilibrium internuclear distance \( R_e \) the ground state molecule is adequately described by the closed-shell HF wave function

$$A[\phi \phi \alpha \beta]$$

(39)

However, in the dissociative limit the system consists of two isolated hydrogen atoms and is accurately described by the open-shell singlet HF wave function

$$A[\phi_1 \phi_2 (\alpha \beta - \beta \alpha)] \quad \langle \phi_1 | \phi_2 \rangle = 0$$

(40)

Since (39) consists of a single doubly occupied orbital, it can never describe the separated-atom limit. Conversely, the required orthogonality between \( \phi_1 \) and \( \phi_2 \) prevents (40) from describing the molecular bond. Hence, while each limit is described by an HF wave function, the inherent restrictions of the HF formalism preclude the possibility of smoothly going from one limit to the other.
The obvious solution of this dilemma is to relax the orthogonality restriction in (40), thereby giving rise to the GVB wave function

\[ A[\phi_1^a \phi_2^b (\alpha \beta - \beta \alpha)] = A[(\phi_1^a \phi_2^b + \phi_2^a \phi_1^b) \alpha \beta] \]  

(41)

where

\[ \langle \phi_1^a | \phi_2^b \rangle = S_{12} \neq 0 \]

Since in (41) no restrictions are placed upon either \( \phi_1^a \) or \( \phi_2^b \), this wave function can behave properly at all internuclear separations with \( S_{12} \) (for the optimum orbitals) varying from zero at infinite separation \([41] = 40\)] to \( \sim 0.8 \) at \( R_e \) [that is, \( (41) = 39\)].\(^{1}\) Furthermore, the added functional freedom associated with this wave function allows for the incorporation of optimal ionic and covalent character in the wave function, thereby leading to a significantly stronger chemical bond than does (39). That is, wave function (41) allows the electrons to better correlate their motions, thus reducing the energy of the system.

For workers conditioned to thinking of chemical systems in terms of traditional molecular HF wave functions it is useful to regard (41) as being a generalization of (39), in which we replace the conventional closed-shell orbital description of a singlet electron pair by a GVB pair consisting of two nonorthogonal orbitals coupled into a singlet,

\[ \phi_1^a \phi_2^a \alpha \beta \rightarrow \phi_1^a \phi_2^a (\alpha \beta - \beta \alpha) \]  

(42a)

or equivalently

\[ \phi_1^a \phi_2^a \alpha \beta \rightarrow (\phi_1^a \phi_2^a + \phi_2^a \phi_1^a) \alpha \beta \]  

(42b)

Using this pair correlation prescription (42), the HF wave function

\[ A[\phi_1^a \phi_2^b \cdots \phi_m^a \alpha \beta \phi_{m+1}^a \alpha \cdots \phi_n^a] \]  

(42c)

would be extended to the wave function of GVB form,

\[ A[\phi_1^a \phi_2^b (\alpha \beta - \beta \alpha) \cdots \phi_1^a \phi_2^b (\alpha \beta - \beta \alpha) \phi_{m+1}^a \alpha \cdots \phi_n^a] \]  

(43)

where

\[ \langle \phi_1^a | \phi_2^b \rangle = S_i \]

In (43) each singlet electron pair is allowed to correlate. However, for most chemical problems many such correlations can be ignored. Since only selected electron pairs need be represented by GVB pairs (e.g., those describing chemical bonds), we normally generalize (42c) to a wave function of the form

\[ A[\{\text{core} \phi_1^a \phi_2^b (\alpha \beta - \beta \alpha) \cdots \phi_1^a \phi_2^b (\alpha \beta - \beta \alpha) \phi_{m+1}^a \alpha \cdots \phi_n^a] \]  

(44)
For example, the excitation energies between the lower electronic states of O₂ are essentially unaffected by correlation effects involving the 1s core orbitals of each oxygen. Indeed, for the lower few states it is not necessary to correlate even the 2s core orbitals of each oxygen.\(^8\)

In wave functions such as (43) or (44) it is computationally useful to require that all orbitals other than the two within a given singlet pair be orthogonal. This strong orthogonality constraint\(^9,10\) is of course a restriction. However, since it applies between orbitals that are normally not expected to overlap significantly, strong orthogonality is usually not a serious restriction.*

The wave function developed here singlet couples as many orbital pairs as is possible for a given eigenstate of spin. Thus for a four-electron singlet state we have

\[ a[φ₁₁ φ₂₁ φ₁₂ φ₂₂(αβ - βα)(αβ - βα)] \]  \hspace{1cm} (45)

However this perfect pairing is not the only coupling scheme possible. For example,

\[ a[φ₁₁ φ₂₁ φ₁₂ φ₂₂[2ααββ + 2ββαα - (αβ + βα)(αβ + βα)]] \]

represents another linearly independent way of coupling these orbitals into a singlet state. The unrestricted GVB wave function\(^1\) allows for these other possibilities as well, having the form

\[ a[φ₁ φ₂ ⋯ φₙ θ(1 ⋯ n)] \]  \hspace{1cm} (46)

where the spin function \(θ\) is allowed full functional freedom within the confines of being an eigenstate of spin

\[ S^2 θ = S(S + 1) θ \]

Furthermore, in the GVB wave function no orbital orthogonality conditions are imposed. The perfect-pairing spin function is especially inappropriate for systems such as the allyl radical\(^11\) or benzene,\(^12\) in which two or more bonding schemes are of comparable importance, and for describing the dissociation of molecules leading to triplet or higher spin states on the separated fragments (e.g., O₂).\(^13\) However, for most cases the orbitals obtained from GVB–PP are quite adequate.

The GVB wave function, in which the perfect pairing and strong orthogonality restrictions are both imposed,\(^3\) is referred to as the GVB–PP (perfect pairing) wave function to distinguish it from the general GVB wave function. In addition, we have developed an intermediate level of calculation in which the spin function is optimized while retaining the strong orthogonality

*These orthogonality conditions have been relaxed in several cases\(^1,2,5\) where it has been found that the orbitals in different singlet pairs of perfect pairing wave functions such as (43) and (44) generally have small overlaps (this arises basically from the Pauli principle).
restrictions. This has been quite useful, for example, for considering allylic systems. These more general GVB wave functions lead to an energy expression not conforming to (6), and hence in this chapter we will restrict ourselves to the GVB–PP approximation.

### 3.2.2. The GVB–PP Energy Expression (The Natural Orbital Representation)

The energy of the simplest GVB wave function (41) is

\[
E = (1 + S_{12}^2)^{-1}[h_{11} + h_{22} + 2S_{12}h_{12} + J_{12} + K_{12}]
\]

(47)

an expression not conforming to (6). However, letting

\[
\phi'_1 = (\sigma_1^{1/2}\phi_1 + \sigma_2^{1/2}\phi_2)/(\sigma_1 + \sigma_2)^{1/2}
\]

\[
\phi'_2 = (\sigma_1^{1/2}\phi_1 - \sigma_2^{1/2}\phi_2)/(\sigma_1 + \sigma_2)^{1/2}
\]

(48)

where \(\sigma_1 > 0, \sigma_2 > 0,\) and \(\langle \phi_1 | \phi_2 \rangle = 0\) allows (41) to be rewritten as

\[
\mathcal{A}\left[(\sigma_1\phi_1 - \sigma_2\phi_2)\alpha\beta\right] = \sigma_1\mathcal{A}[\phi_1\alpha\beta] - \sigma_2\mathcal{A}[\phi_2\alpha\beta]
\]

(49)

giving rise to the energy expression

\[
E = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(2h_{11} + J_{11}) + \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}(2h_{22} + J_{22}) - \frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}K_{12}
\]

(50)

Equation (50) does have the desired form of (6), differing from a typical HF expression only in that the energy coefficients \((f, a, b)\) are now functions of the pair coefficients \(\sigma_1\) and \(\sigma_2\), rather than being fixed.

Thus, while it is conceptually the nonorthogonal GVB orbitals that are the most useful, computationally it is much more convenient to replace the GVB pair,

\[
\phi'_1, \phi'_2(\alpha\beta - \beta\alpha), \quad \langle \phi'_1 | \phi'_2 \rangle = S_{1i}
\]

(51)

with its natural orbital\(^{(14)}\) representation,

\[
(\sigma_1\phi_1 - \sigma_2\phi_2)\alpha\beta, \quad \langle \phi_1 | \phi_{2i} \rangle = 0
\]

(52)

where

\[
\phi'_1 = (\sigma_1^{1/2}\phi_1 + \sigma_2^{1/2}\phi_2)/(\sigma_1 + \sigma_2)^{1/2}
\]

\[
\phi'_2 = (\sigma_1^{1/2}\phi_1 - \sigma_2^{1/2}\phi_2)/(\sigma_1 + \sigma_2)^{1/2}
\]

(53)

\[
\langle \phi_1 | \phi_{2i} \rangle = 0, \quad \sigma_{ii} > 0
\]

and where for future convenience we take

\[
\sigma_{1i}^2 + \sigma_{2i}^2 = 1
\]

(54)
Once written in terms of these natural orbital pairs, the energy of any GVB–PP wave function always has the form of (6). For example, casting the two-pair wave function,

$$a[(\text{core})\phi'_1\phi'_2\phi'_3\phi'_4(\alpha\beta - \beta\alpha)(\alpha\beta - \beta\alpha)]$$  \hspace{1cm} (55)

in its natural orbital form

$$a[(\text{core})(\sigma_1\phi_1 - \sigma_2\phi_2)(\sigma_3\phi_3 - \sigma_4\phi_4)\alpha\beta\alpha\beta]$$  \hspace{1cm} (56)

leads to

$$E = E_c + \sum_{i=1}^{4} \sigma_i^2(2\hbar_i + J_i) + 2 \sum_{i=1}^{2} \sum_{j=3}^{4} \sigma_i^2\sigma_j^2(2J_{ij} - K_{ij})$$

$$-2\sigma_1\sigma_2K_{12} - 2\sigma_3\sigma_4K_{34}$$  \hspace{1cm} (57)

Thus, the GVB–PP wave function (in natural orbital form) having the general form

$$a[(\text{core})\{\text{pair}\}\{\text{open}\}]$$  \hspace{1cm} (58)

where*

\{core\} = \phi_{c_1}\phi_{c_2}\alpha\beta\phi_{c_2}\phi_{c_2}\alpha\beta \cdots

\{pair\} = (\sigma_{p_1}\phi_{p_1}\phi_{p_1} - \sigma_{p_2}\phi_{p_2}\phi_{p_2})\alpha\beta(\sigma_{p_3}\phi_{p_3}\phi_{p_3} - \sigma_{p_4}\phi_{p_4}\phi_{p_4})\alpha\beta \cdots

\{open\} = \phi_{o}\phi_{i+1}\alpha \cdots

leads to an energy expression in the desired form (6), where

$$f_i = 1 \quad \text{if } \phi_i \text{ is a core orbital}$$

$$f_i = \sigma_i^2 \quad \text{if } \phi_i \text{ is a pair orbital}$$

$$f_i = \frac{1}{2} \quad \text{if } \phi_i \text{ is an open orbital}$$  \hspace{1cm} (59)

$$a_{ii} = 2f_if_i$$

$$b_{ii} = -f_if_i$$

except that

$$b_{ij} = -\frac{1}{2} \quad \text{if } \phi_i \text{ and } \phi_j \text{ are both open orbitals}$$

$$a_{ii} = f_i$$

$$b_{ii} = 0$$  \hspace{1cm} \{\text{if } \phi_i \text{ is a pair orbital}\}

$$a_{ii} = 0$$

$$b_{ii} = -\sigma_i\sigma_i$$  \hspace{1cm} \{\text{if } \phi_i \text{ and } \phi_j \text{ are in the same pair}\}

*The \{open\} wave function can also be taken of the form in (33) or as a product of (33) with a multiplet open shell. The modifications of the coefficients (59) are straightforward.
Upon solving for the GVB natural orbitals \( \phi_{1i} \) and \( \phi_{2i} \) and the pair coefficients \( \sigma_{1i} \) and \( \sigma_{2i} \), the GVB orbitals \( \phi'_1i \) and \( \phi'_2i \) are obtained from (48), where the overlap \( S_i \) is given by

\[
S_i = \frac{\sigma_{1i} - \sigma_{2i}}{\sigma_{1i} + \sigma_{2i}}
\]  

(60)

3.3. Separated-Pair Wave Functions

Written in its natural orbital form, we see that another prescription for obtaining a GVB-PP wave function from the usual HF wave function is

\[
\phi_\rho \phi_\sigma \alpha \beta \to (\sigma_{1p} \phi_{1p} \phi_{1p} - \sigma_{2p} \phi_{2p} \phi_{2p}) \alpha \beta, \quad \langle \phi_{1p} | \phi_{2p} \rangle = 0
\]  

(60a)

That is, an electron pair normally described by a closed-shell orbital in the HF wave function is instead described by a geminal expansion consisting of two orthogonal doubly occupied orbitals. In a separated-pair wave function\(^{(15,16)}\) this pair correlation prescription is extended to the general form,*

\[
\phi_\rho \phi_\sigma \alpha \beta \to \sum_{i=1}^{n_p} \sigma_{ip} \phi_{ip} \phi_{ip} \alpha \beta, \quad \langle \phi_{ip} | \phi_{ip} \rangle = \delta_{ii}
\]  

(60b)

so that the pair functional consists of an expansion of arbitrary length. That these orbitals can be taken as orthogonal without restriction becomes obvious upon realizing that the general singlet functional,

\[
\sum_{i,j} d_{ij} \phi'_{ip} \phi'_{ip} \alpha \beta
\]

where

\[
\langle \phi'_{ip} | \phi'_{ip} \rangle = \delta_{ii}, \quad d_{ij} = d_{ji}
\]

is directly related to (60b) through a general orthogonal transformation of the form

\[
\phi'_i = (\phi_{ip} + \lambda_i \phi_{ip})/(1 + \lambda_i^2)^{1/2}, \quad \lambda_{ij} = -\lambda_{ji} \quad \text{for all } i \text{ and } j
\]

Thus, a separated-pair functional can, in principle, be used to introduce any amount of correlation between the two singlet coupled electrons by taking \( n_p \) to be arbitrarily large. In practice, all significant correlation effects can normally be taken into account using a relatively small expansion through an optimal choice of the orbitals involved.\(^{(15,16)}\)

*Note that for two-term wave functions we define a minus sign with \( \sigma_2 \) so that \( \sigma_1, \sigma_2 > 0 \). For separated-pair wave functions we let all terms enter with + \( \sigma_i \) coefficient so that some \( \sigma_i < 0 \).
The ability of (60b) to correlate the two electrons within a given singlet pair to an arbitrary degree can at times be quite important and can obviously lead to results that are quantitatively superior to those possible using the GVB pair functional (60a). However, for most chemical problems the dominant correlations (both qualitative and quantitative) can usually be accomodated by the simpler GVB pair functional. Additional effects, which could be incorporated through use of (60b) with $n_p > 2$, are generally of less importance, especially when compared with other types of correlations possible in a many-electron system.

The GVB wave function (46) obtained by relaxing the perfect-pairing and strong orthogonality constraints associated with a GVB–PF wave function can incorporate many of these additional kinds of correlations as well, and we regard the GVB–PP wave function as a restricted GVB wave function, not as a restricted separated-pair wave function. This distinction, which on the basis of comparing (60a) and (60b) is rather obscure, crystallizes upon recalling that the GVB pair functional in (60a) is simply a computationally convenient representation for a pair of nonorthogonal orbitals coupled into a singlet. Therefore, just as it is the case for HF and GVB wave functions, a GVB–PP wave function is readily susceptible to physical interpretation through the fundamental conceptualization of associating each electron with a special orbital. On the other hand, for a separated-pair wave function ($n_p > 2$) this is not possible since each correlated electron pair is described in terms of more than two orbitals.

On a computational level, however, there really is no formal distinction. That is, using the conventional normalization criterion

$$\sum_{i=1}^{n} \sigma_{ip}^2 = 1$$

the energy expression and actual energy coefficients as given in (59) are the same for both GVB–PP and separated-pair wave functions and hybrids thereof. The only difference is that now a "pair" can be composed of more than two orbitals. This being the case, we find it convenient to refer to such wave functions collectively as correlated-pair wave functions and the orbitals involved in such pairs are then simply called correlated-pair orbitals.

4. The Basic Variational Equations

Having demonstrated that energy expression (6) is applicable to a wide range of useful wave functions, we will now concern ourselves with the task of solving for the optimal (self-consistent) orbitals of (6); generally those leading to the lowest energy. Since we will be dealing solely with (6), where the energy coefficients $\{f_i, a_{ip}, b_{ij}\}$ are regarded as arbitrary parameters, the results obtained are quite general in that they apply to any wave function (actual or
contrived) involving orthonormal orbitals whose energy conforms to (6).
Moreover, our underlying approach is not rooted in the simple homogeneous
form of (6) and can therefore be readily extended to the more general case.

In solving this problem, we will use the variational principle as our guide.
That is, we will seek those orbitals \( \{ \phi_i \} \) for which the energy is stationary (at a
minimum) with respect to any arbitrary allowed variations in any of the
orbitals. For convenience of notation we will take all orbitals as real. Extension
to complex orbitals is straightforward.

4.1. The Variational Equation and Fock Operators

Starting with the optimal orbitals \( \{ \phi_i \} \) for (6) having energy \( E \), consider a
slightly modified set of orbitals \( \{ \phi'_i \} \) having energy

\[
E' = \sum_i 2f_i h_{\phi_i} + \sum_{ij} (a_{ij} J_{\phi_i, \phi_j} + b_{ij} K_{\phi_i, \phi_j})
\]

(61)

Since the form of (61) is applicable only for orthonormal orbitals, we must
require that this new set satisfy

\[
\langle \phi'_i | \phi'_j \rangle = \delta_{ij}
\]

(62)

Recognizing that the only meaningful change in an orbital is one which is
orthogonal to that orbital, we can write the modified orbital \( \phi'_i \) in terms of the
optimal orbital \( \phi_i \) as

\[
\phi'_i = (\phi_i + \delta\phi_i)/(1 + \langle \delta\phi_i | \delta\phi_i \rangle)^{1/2}
\]

(63a)

where

\[
\langle \phi_i | \delta\phi_i \rangle = 0
\]

(63b)

and \( \delta\phi_i \) is the variation in this orbital. Expanding \( E' \) in terms of \( \{ \phi_i \} \) and \( \{ \delta\phi_i \} \)
yields

\[
E' = E + \delta^{(1)} E + \delta^{(2)} E + \cdots
\]

(64)

where the first-order variation in the energy is

\[
\delta^{(1)} E = 4 \sum_i \langle \delta\phi_i | \left( f_i h + \sum_j (a_{ij} J_j + b_{ij} K_j) \right) | \phi_i \rangle
\]

(65)

Defining the generalized Fock operator,

\[
F_i = f_i h + \sum_j (a_{ij} J_j + b_{ij} K_j)
\]

(66)

allows (65) to be written as

\[
\delta^{(1)} E = 4 \sum_i \langle \delta\phi_i | F_i | \phi_i \rangle
\]
Since the energy $E$ is at a minimum with respect to changes in the orbitals, it must then be stationary through first-order variations in these orbitals. That is,

$$\delta^{(1)} E = 0$$

(67)

if the $\{\phi_i\}$ are indeed optimal. Hence, a necessary condition for $\{\phi_i\}$ to be optimal is that

$$\sum_i \langle \delta \phi_i | F_i | \phi_i \rangle = 0$$

(68)

for all orbital variations consistent with (62) and (63).

It is important to keep in mind that this result (68) with $F_i$ in the form of (66) is valid only when the orthogonality condition (62) is maintained (at least through first order).

Having defined the Fock operators (66), we find that the energy can be written as

$$E = \sum_i [f_i h_i + \langle \phi_i | F_i | \phi_i \rangle]$$

(69a)

From this expression we see that these operators are somewhat ambiguous in that,

$$F'_i = F_i + g_i (J_i - K_i)$$

(69b)

(where $g_i$ is a completely arbitrary parameter) is an equally valid Fock operator for orbital $i$. This result is of course simply a reflection of (15).

**Fock Operator Examples (Shells)**

Using (66) we find from (18) that the Fock operator for the orbitals of closed-shell wave function (17) is

$$F_c = h + \sum_i (2J_i - K_i)$$

(70a)

for all orbitals. That is, we obtain the same operator for each closed-shell orbital. Attainment of this result was the main reason for adding and subtracting self-terms in the energy expression for such wave functions. Had we neglected to do so, (66) would have led to a different operator for each orbital,

$$F_i = h + J_i + \sum_{i \neq i} (2J_i - K_i)$$

(70b)

However, we could also arrive at (70a) by using (69b) and choosing $g_i = 1$. Extending to wave functions involving open-shell orbitals as well, the closed-shell Fock operator becomes

$$F_c = h + \sum_i f_i (2J_i - K_i)$$

(71)
For high-spin (multiplied) coupled open-shell orbitals, as in wave function (31), we obtain

\[ 2F_m = h^e + \sum J_i - K_i \]  

(72)

Thus, all the Fock operators for multiplet coupled orbitals can be the same. This remains true even if additional open-shell orbitals are present such as in (36) and (55). Note, however, that

\[ F_c \neq F_m \]  

(73)

so that even for the simplest wave functions involving both closed-shell and open-shell orbitals there is no single Fock operator applicable to all orbitals.

In most other instances each orbital has its own distinct Fock operator. For example, for the open-shell singlet wave function (33) we obtain

\[ 2F_1 = h^e + J_2 + K_2 \]  

(74)
\[ 2F_2 = h^e + J_1 + K_1 \]

and there is no way of rewriting the energy expression [or choosing \( g_1 \) and \( g_2 \) in (69b)] to obtain the same operator for each orbital. Likewise, in a correlated-pair wave function, the Fock operator for the \( k \)th orbital of the \( l \)th correlated pair,

\[ F_{kl} = \sigma_{kl}^2 \left[ h^e + J_{kl} + \sum_{j_1, j_2, \ldots} f_j (2J_j - K_j) \right] + \sum_{i \neq k} \sigma_{kl} \sigma_{il} K_{il} \]  

(75)

is unique to this orbital.

By convention, orbitals which can have the same Fock operator are said to belong to the same shell. From the preceding examples we see that the closed and multiplet shells may contain several orbitals. On the other hand, the orbitals in a correlated-pair or HF open-shell singlet pair each belong to a unique shell. If a shell is capable of multiple occupancy, the distinct Fock operator applicable to all of the orbitals in the shell is referred to as the Fock operator for that shell.* For a shell containing only one orbital, the Fock operator for that shell is generally taken as the one containing no self-exchange term (\( b_{\nu \nu} = 0 \)).

4.2. The Variational Conditions

Since Eq. (68) must be satisfied for all allowed orbital variations, we can explore its implications by focusing attention on all possible changes in a particular orbital, say \( \phi_{\nu} \). In solving for the wave functions of molecules we will

*Since the \( F \) operator for such a shell is symmetric with respect to the orbitals of that shell, it is invariant to any orthogonal (unitary) transformation among these orbitals. Therefore, from (69a) it follows that the energy is also invariant to such transformations between these orbitals.
exclusively consider the basis set expansion approach. That is, all orbitals and their variations will be expanded in terms of a set of \( P \) linearly independent basis functions, thus defining the \( P \)-dimensional Hilbert space in terms of which the system is to be described. In the following we will make a distinction between the occupied orbitals, that is the \( n \) orbitals involved in (6) (at any stage in the iterative process), and the \( P-n \) virtual orbitals, also constructable in terms of the \( P \)-dimension basis being employed.

Allowing only orbital \( \phi_v \) to vary, (68) reduces to

\[
\langle \delta \phi_v | F_v | \phi_v \rangle = 0
\] (76)

However, from (62) and (63), for \( j \neq \nu \), we have

\[
0 = \langle \phi_j | \phi'_\nu \rangle = \frac{\langle \phi_j | \phi_v + \delta \phi_v \rangle}{(1 + \langle \delta \phi_v | \delta \phi_v \rangle)^{1/2}}
\] (77)

where

\[
\langle \phi_i | \phi_j \rangle = \delta_{ij}, \quad \langle \delta \phi_v | \phi_v \rangle = 0
\]

Thus

\[
\langle \phi_j | \delta \phi_v \rangle = 0 \quad \text{for all } j
\] (78)

That is, only those variations which are orthogonal to all orbitals may be considered. Therefore (76) provides us with a necessary condition for \( \phi_v \) to be optimum with respect to any variations orthogonal to all occupied orbitals.

Since (76) applies only to variations orthogonal to all occupied orbitals, we must still consider the case where orbital \( \nu \) is allowed to vary within the space of other occupied orbitals. For example, consider the case where \( \phi'_\nu \) is allowed to overlap with \( \phi_\mu \),

\[
\phi'_\nu = (\phi_v + \lambda \phi_\mu)/(1 + \lambda^2)^{1/2}
\] (79)

so that

\[
\delta \phi_v = \lambda \phi_\mu
\] (80)

Since orbital orthogonality must be preserved, orbital \( \mu \) must change simultaneously and must satisfy

\[
\delta \phi_\mu = -\lambda \phi_v
\] (81)

That is, \( \phi'_\nu \), as given in (79) dictates that \( \phi'_\mu \) must have the form

\[
\phi'_\mu = (\phi_\mu - \lambda \phi_v)/(1 + \lambda^2)^{1/2}
\] (82)
Allowing no additional variations, Eq. (68) reduces to
\[
0 = \langle \delta \phi_\nu | F_\nu | \phi_\nu \rangle + \langle \delta \phi_\mu | F_\mu | \phi_\mu \rangle \\
= \lambda \langle \phi_\nu | (F_\nu - F_\mu) | \phi_\mu \rangle
\]
(83)
from which we obtain
\[
\langle \phi_\nu | (F_\nu - F_\mu) | \phi_\mu \rangle = 0
\]
(84)
as a necessary condition for orbitals \( \phi_\nu \) and \( \phi_\mu \) to be optimum with respect to each other.*

Condition (76) ensures that \( \delta^{(1)} E = 0 \) for any variation \( \delta \phi_\nu \) (in orbital \( \phi_\nu \)) that is orthogonal to all occupied orbitals, and condition (84) (for all \( \mu \)) ensures that \( \delta^{(1)} E = 0 \) for any variation \( \delta \phi_\nu \), overlapping any occupied orbital. Thus conditions (76), for all \( \nu \), and (84), for all \( \mu \) and \( \nu \), are sufficient (and necessary) to ensure that \( \delta^{(1)} E = 0 \) for all variations of all orbitals. Together these equations are referred to as the variational conditions.

4.3. The Orbital Correction Equation

Although the variational conditions determine whether a given set of orbitals is optimal, they do not tell us how to improve upon a nonoptimal set. That is, they do not tell us how to solve for an optimal set of orbitals. We will now consider that we have some initial guess \( \{ \phi_i^0 \} \) for the orbitals and that we want to find the corrections \( \{ \Delta \phi_i \} \) leading to the optimum orbitals \( \{ \phi_i \} \). In order to do this we return to the energy expression and re-expand, retaining all terms through second order in \( \{ \Delta \phi_i \} \). Applying the variational principle to the resulting expression then leads to equations that are valid through first order† in \( \{ \Delta \phi_i \} \) and that can be solved to obtain the optimal corrections.

Since orbital orthonormality must always be preserved, these corrections cannot be fully independent of each other. Expanding each orbital as
\[
\phi_i = (\phi_i^0 + \Delta \phi_i) / (1 + \langle \Delta \phi_i | \Delta \phi_i \rangle)^{1/2}
\]
(85)
where
\[
\langle \phi_i^0 | \Delta \phi_i \rangle = 0
\]
ensures that
\[
\langle \phi_i | \phi_i \rangle = 1
\]

*Note that if orbitals \( \mu \) and \( \nu \) are in the same shell \( (F_\mu = F_\nu) \), then (84) is always satisfied.

†We will also retain some higher-order terms (vide infra).
Since \( \{ \phi_i^0 \} \) and \( \{ \phi_i \} \) each constitute an orthogonal set, we obtain
\[
0 = \langle \phi_i | \phi_j \rangle
\]
\[
= \frac{\langle \Delta \phi_i | \phi_j^0 \rangle + \langle \phi_i^0 | \Delta \phi_j \rangle + \langle \Delta \phi_i | \Delta \phi_j \rangle}{(1 + \langle \Delta \phi_i | \Delta \phi_j \rangle)^{1/2}(1 + \langle \Delta \phi_i | \Delta \phi_j \rangle)^{1/2}} \quad i \neq j \tag{86}
\]

Thus, \( \{ \Delta \phi_i \} \) must be constrained to satisfy
\[
\langle \Delta \phi_i | \phi_j^0 \rangle + \langle \phi_i^0 | \Delta \phi_j \rangle + \langle \Delta \phi_i | \Delta \phi_j \rangle = 0 \quad i \neq j \tag{87}
\]

A logical starting point for considering a variational determination of \( \{ \Delta \phi_i \} \) is the expression for the change in energy due to these corrections,
\[
\Delta E = E - E^0
\]
\[
= 2 \sum_i f_i (h_{ii} - h_{i0i0}) + \sum_i \sum_{i,j} [a_{ij} (J_{ij} - J_{i0j0}) + b_{ij} (K_{ij} - K_{i0j0})] \tag{88}
\]

Substituting (85) into this expression and collecting terms yields
\[
\frac{1}{2} \Delta E = \sum_i (1 + \langle \Delta i | \Delta i \rangle)^{-1} f_i [2 \langle \Delta i | h | i_0 \rangle + \langle \Delta i | h | \Delta i \rangle - \langle \Delta i | \Delta i \rangle h_{i0i0}]
\]
\[
+ \sum_i (1 + \langle \Delta j | \Delta j \rangle)^{-1} [a_{ij} (2 \langle \Delta j | J_{j0} | i_0 \rangle + \langle \Delta j | J_{j0} | \Delta i \rangle - \langle \Delta j | \Delta i \rangle J_{j0i0})
\]
\[
+ 2 (i_0 \Delta j | J_{j0} | i_0) + 2 \langle \Delta j | J_{j0} | \Delta i \rangle - \frac{1}{2} \langle \Delta i | \Delta i \rangle J_{j0i0} - \frac{1}{2} \langle \Delta j | \Delta j \rangle J_{j0i0})
\]
\[
+ b_{ij} (2 \langle \Delta j | K_{j0} | i_0 \rangle + \langle \Delta j | K_{j0} | \Delta i \rangle - \langle \Delta i | \Delta i \rangle K_{j0i0} + \langle \Delta i | \Delta j | i_0 \rangle)
\]
\[
+ (i_0 \Delta j | J_{j0} | i_0) + 2 \langle \Delta j | K_{j0} | i_0 \rangle + \frac{1}{2} K_{j0i0} - \frac{1}{2} \langle \Delta i | \Delta i \rangle \langle \Delta j | \Delta j \rangle K_{j0i0})] \tag{89}
\]

where we have used the shorthand notation
\[
i = \phi_i, \quad i_0 = \phi_i^0, \quad \Delta i = \Delta \phi_i
\]

Bear in mind here that (89) is subject to the constraint of (87).

In principle, \( \{ \Delta \phi_i \} \) could be variationally obtained from (89) and (87) exactly. However, in practice this would eventually require us to solve for the (real) solutions for an extremely complicated set of coupled high-order equations. The alternative is to consider an iterative procedure for obtaining \( \{ \phi_i \} \). In this way, during each iteration we need only seek corrections leading to better orbitals for use in the next iteration. Since only approximate solutions for \( \{ \Delta \phi_i \} \) will now be needed, we can base such a procedure on an approximation to (89) in which troublesome high-order terms are neglected. We will now consider such an approximation.

Rewriting (89) in terms of the zero-order Fock operators
\[
\hat{F}_{i0} = f_i h + \sum_j (a_{ij} J_{jo} + b_{ij} K_{jo}) \tag{90}
\]
we obtain
\[
\Delta E = \sum_i \left( 1 + (\Delta i|\Delta i) \right)^{-1} \left\{ 2(\Delta i|F_{i0}|i_0) + (\Delta i|F_{i0}|\Delta i) - (\Delta i|i_0|F_{i0}|i_0) \right\} \\
+ \sum_i \left[ 2a_{ij}(i_0\Delta i|j_0\Delta j) + b_{ij}(\Delta i\Delta j|i_0j_0) + b_{ij}(i_0\Delta j|j_0\Delta i) \right] + O(\Delta^3) \quad (91)
\]
where \(O(\Delta^3)\) contains only terms third order and higher in \(\{\Delta \phi_i\}\) and is of the form
\[
O(\Delta^3) = \sum_{i,j} \left( 1 + (\Delta i|\Delta i) \right)^{-1}(1 + (\Delta j|\Delta j))^{-1} O_{ij}(\Delta^3)
\]

Defining \(\Delta F_{i0}(j_0 \to \Delta j)\) as the first-order correction to \(F_{i0}\) due to the change \(\Delta \phi_j\) in orbital \(j\), the total first-order correction is
\[
\Delta F_{i0} = \sum_j \Delta F_{i0}(j_0 \to \Delta j) \quad (92)
\]
and we find that
\[
(\Delta i|\Delta F_{i0}|i_0) = \sum_j (\Delta i|\Delta F_{i0}(j \to \Delta j)|i_0)
\]
\[
= \sum_j \left[ 2a_{ij}(i_0\Delta i|j_0\Delta j) + b_{ij}(\Delta i\Delta j|i_0j_0) + b_{ij}(i_0\Delta j|j_0\Delta i) \right] \quad (93)
\]

Therefore, defining \(\Delta \varepsilon\) as the part of \(\frac{1}{2} \Delta E\) containing terms up through order \(\Delta^2\), we obtain the result
\[
\Delta \varepsilon = \sum_i \left( 1 + (\Delta i|\Delta i) \right)^{-1} \left\{ 2(\Delta i|F_{i0}|i_0) + (\Delta i|F_{i0}|\Delta i) \right\} \\
+ (\Delta i|\Delta F_{i0}|i_0) - (\Delta i|i_0|F_{i0}|i_0) \quad (94a)
\]

Alternatively, we can define a total \(\varepsilon\) as
\[
\varepsilon = \Delta \varepsilon + \sum_i (i_0|F_{i0}|i_0)
\]
\[
= E^0 + \sum_{ij} (a_{ij}J_{i0} + b_{ij}K_{i0})
\]
leading to
\[
\varepsilon = \sum_i \left( 1 + (\Delta i|\Delta i) \right)^{-1} \left\{ (i_0 + \Delta i)|F_{i0}|(i_0 + \Delta i) + (\Delta i|\Delta F_{i0}|i_0) \right\} \quad (94b)
\]
In deriving (94a, b) we have purposely retained individual orbital renormalization terms. In this way we impart to Δε and ε the essential functional dependence upon {Δφi} consistent with preservation of orbital normalization. However, we have made no attempt to incorporate orbital orthogonality constraints. Consequently, in variationally determining {Δφi} using (94) we must still explicitly require that {Δφi} satisfy

\[ \langle i_0 | Δφ_i \rangle + \langle j_0 | Δφ_j \rangle + \langle Δφ_i | Δφ_j \rangle = 0 \quad i \neq j \]

\[ \langle i_0 | Δφ_i \rangle = 0 \]  

Equations (94) and (95) constitute the basic orbital correction equations which we must solve to obtain corrections leading to new and better orbitals,

\[ φ_i^1 = (φ_i^0 + Δφ_i) / (1 + Δφ_i | Δφ_i) )^{1/2} \]  

These orbitals {φi^1} will then serve as the initial guesses for another iteration until self-consistency is achieved. Since all terms through second order are retained in (94), these correction terms will be valid through first order and hence the iterative solution for the optimum orbitals {φi} should converge quadratically. Substituting (95) into (94) and requiring that \( \partial Δε / \partial Δi = 0 \) for all \( i \) leads to a set of coupled equations [see (127) below] that must be solved to obtain the optimal corrections \( {Δφ_i} \). For a system with \( P \) basis functions and \( n \) occupied orbitals, there are \( nP \) such equations determining the optimal corrections, the formation and solution of which is currently impractical for cases of most interest. As a result, in the following section we will partition the problem into several more restricted variations, each of which will be sequentially applied.

5. The Iterative Self-Consistent Field Equations

Rigorously applied, Eqs. (94) and (95) permit simultaneous solution for all Δφi. This straightforward approach ultimately requires (for the case of \( n \) orbitals and \( P \) basis functions) simultaneous solution of at least \( nP \) coupled nonlinear equations. However, by appropriately decoupling these equations the computational problem can be greatly simplified while retaining adequate convergence. Thus, we partition each iterative cycle into two separate steps.

a. Optimization of each orbital with respect to the space orthogonal to all orbitals, hereafter referred to as the virtual space. This step utilizes an orthogonality constrained basis set expansion (OCBSE) and is referred to as the OCBSE step.
b. Optimization of the orbitals with respect to each other. This step is referred to as orbital mixing.

5.1. OCBSE Optimization

Let us first consider optimization of just one orbital ($\phi_\nu$) while keeping all other orbitals fixed. Thus, (94b) reduces to

$$
\varepsilon_\nu = (1 + \langle \Delta \nu | \Delta \nu \rangle)^{-1} \left[ \langle \nu_0 + \Delta \nu | F_{\nu 0} | \nu_0 + \Delta \nu \rangle + \langle \Delta \nu | \Delta F_{\nu 0} (\nu_0 \rightarrow \Delta \nu) | \nu_0 \rangle \right] \tag{97}
$$

where from (95) the change $\Delta \phi_\nu$ in orbital $\nu$ must satisfy

$$
\langle i_0 | \Delta \phi_\nu \rangle = 0 \quad \text{for all } i \tag{98a}
$$

Hence, $\Delta \phi_\nu$ can only involve the virtual space, that is the space orthogonal to all occupied orbitals. Evaluating the Fock correction term in (97), we obtain

$$
\varepsilon_\nu = (1 + \langle \Delta \nu | \Delta \nu \rangle)^{-1} 
\times \left[ \langle \nu_0 + \Delta \nu | F_{\nu 0} | \nu_0 + \Delta \mu \rangle + (2a_{\nu \nu} + b_{\nu \nu})K_{\nu 0 \Delta \nu} + b_{rr}J_{\nu 0 \Delta \nu} \right] \tag{98b}
$$

However, since the operator

$$
F_{\nu} = F_{\nu} + b_{\nu \nu} (J_{\nu} - K_{\nu}) 
= f_{\nu} h + \sum_{j=\nu} (a_{\nu j} J_{j} + b_{\nu j} K_{j}) + (a_{\nu \nu} + b_{\nu \nu}) J_{\nu} \tag{98c}
$$

is equally valid for this orbital, we can work with the simpler yet equivalent expression

$$
\varepsilon_\nu = (1 + \langle \Delta \nu | \Delta \nu \rangle)^{-1} \left[ \langle \nu_0 + \Delta \nu | F_{\nu}^* | \nu_0 + \Delta \nu \rangle + 2a_{\nu \nu} K_{\nu 0 \Delta \nu} \right] \tag{98d}
$$

where

$$
a'_{\nu \nu} = a_{\nu \nu} + b_{\nu \nu}
$$

As usual, we solve for the orbitals in terms of a finite basis set,

$$
\phi_i = \sum_{i=1}^{P} d_i \chi'_i \tag{99}
$$

The primitive basis functions $\{\chi'_i\}$ used in such expansions are generally of a type which facilitate evaluation of molecular integrals (e.g., atomic-like functions of Slater or Gaussian form). However, during the iterative SCF procedure
it is convenient to redefine the basis (in terms of these primitives) as an
orthonormal set consisting of the $n$ current orbitals $\{\phi^0_\nu\}$ plus $P-n$ virtuals, $\{\chi_i\}$. In this way we can automatically satisfy (98) by explicitly excluding the other
occupied orbitals $\{\phi^0_i, i \neq \nu\}$ from the expansion for $\phi_\nu$.

$$
\phi_\nu = \left( \phi^0_\nu + \sum_{i=1}^{P-n} d_{\nu i} \chi_i \right) / \left( 1 + \sum_i d_{\nu i} \right)^{1/2}
$$

(100)

so that

$$
\Delta \phi_\nu = \sum_{i=1}^{P-n} d_{\nu i} \chi_i
$$

(101)

This is referred to as the orthogonality constrained basis set expansion, or
OCBSE.

Substituting (101) in (98d) we obtain

$$
\epsilon_\nu = \left( 1 + \sum_i d_{\nu i}^2 \right)^{-1} \left[ (\nu_0|F_{\nu 0}|\nu_0) + 2 \sum_i d_{\nu i} (\chi_i|F_{\nu 0}|\nu_0) + \sum_{i,j} d_{\nu i} d_{\nu j} (\chi_i|(F_{\nu 0} + 2a_{\nu i} K_{\nu 0})|\chi_j) \right]
$$

(102)

If we define

$$
\chi_0 = \phi^0_\nu
$$

and let

$$
C_{\nu i} = d_{\nu i} d_{0 \nu}
$$

(103)

then (102) becomes

$$
\sum_{i=1}^{P-n} C_{\nu i} C_{\nu i} (\bar{F}_{\nu 0})_{ii} = \epsilon_\nu \sum_{i=0}^{P-n} C_{\nu i}^2
$$

(104)

where in matrix notation

$$
\bar{F}_{\nu 0} = F_{\nu 0} + X_{\nu 0}
$$

(105a)

and

$$
(X'_{\nu 0})_{ii} = 2a_{\nu i} (1 - \delta_{i,0})(1 - \delta_{i,0})(\chi_i|K_{\nu 0}|\chi_i)
$$

(106a)

The variation equation from (104) is

$$
\bar{F}_{\nu 0} C_\nu = C_\nu \epsilon_\nu
$$

(105b)

Thus, with OCBSE the optimum orbital $\phi_\nu$ is obtained by merely diagonalizing
the $\bar{F}_{\nu 0}$ matrix and selecting the appropriate root. Usually we want the orbital
leading to the lowest total energy and, therefore, we select the root having the
most negative eigenvalue. We find that even when there are large changes in
the orbital (105b) generally has a sufficiently wide radius of convergence to accommodate such changes. In certain circumstances we can deliberately force convergence to self-consistent wave functions describing excited states by choosing appropriate higher roots [e.g., by selecting from the roots of (105b) the one that has the highest overlap with $\phi^{0}_r$].

A common approach to variational equations has been to take the *first-order* variational equation\(^{(17)}\)

$$F'_rC_v = \varepsilon_vC_v$$

that must be satisfied by the optimum orbital $\phi_v$, replace the operator $F'_r$ with $F'_{m_r}$, and solve

$$F'_{m_r}C_v = \varepsilon_vC_v \quad (106b)$$

to obtain a new orbital. Comparing with (105b) we see that this procedure is correct if there are no self-terms, that is, is $a_{\nu\nu} = b_{\nu\nu} = 0$ in (98c). Otherwise (106b) should include the terms (106a) and as a result (106b) is expected to have a much smaller radius of convergence than (105b).

By virtue of both the $F'$ and $X'$ matrices, the $\tilde{F}$ matrix in (105a) is different for each orbital of a shell. However, letting orbital $\nu$ belong to shell $s$ we can rewrite (105a) in terms of the Fock operator $F_s$ for the shell $s$ to give

$$(F_{s_0} + X_{s_0}^S)C_v = C_s\varepsilon_v \quad (107a)$$

where

$$(X_{s_0}^S)_{ii} = (1 - \delta_{i,0})(1 - \delta_{i,0})\langle \chi_i | [[(2a_{ss} + b_{ss})K_{ss} + b_{ss}J_{ss}]\chi_i \rangle \quad (107a)$$

In this form, the $\tilde{F}$ matrices for each orbital in a given shell differ only by virtue of the $X$ matrices. If orbital $\nu$ is self-consistent,* this fact is signaled by

$$\langle \chi_i | F_s \phi_{s_0} \rangle = 0 \quad i = \text{all virtuals} \quad (107c)$$

That is, the matrix $X_{s_0}^S$ is not needed for verification of having attained self-consistency for orbital $\nu$. The sole role of this matrix is to provide the rigorous single iteration redefinition of an as yet unconverged orbital. In order to attain the computation advantages of solving for a whole shell of orbitals with one operator, we can neglect this *self-correction* matrix and take our OCBSE optimization equation to be simply

$$F_{s_0}C_v = C_s\varepsilon_v \quad (108a)$$

This approximation generally leads to sufficiently rapid convergence and is usually used in place of (107a) without adverse effect. If orbital $\nu$ belongs to its own unique shell, there is little to be gained by this approximation and we should use the more correct expression (105a).

*Within this context the "self-consistency" of orbital $\nu$ is with respect to the fixed field due to $\phi_{s_0}^i$; $i \neq \nu$. 
The major advantage in using (108a) occurs in cases where several orbitals belong to the same shell. For example, for wave functions involving any number of closed-shell orbitals (108a) leads to

\[
F_c C_i = C_i \varepsilon_i \quad \text{\( i = \text{all closed-shell orbitals} \)} \quad (108b)
\]

Therefore, since each orbital corresponds to a different (orthogonal) solution of the same eigenvalue equation, we can simultaneously obtain new guesses for all these orbitals by simply constructing and diagonalizing a single \(F_c\) matrix (constructed over the OCBSE basis excluding any current open-shell orbitals). This contrasts with the case of the more rigorous equations (105a) or (107a) which lead to

\[
\tilde{F}_i \neq \tilde{F}_j
\]

and hence to different eigenvalue problems for each closed-shell orbital. In such instances the significant computational simplification obtained using (108) is expected to usually outweigh the advantages of a somewhat increased convergence which could result from (107a).

5.2. Orbital Mixing

Thus far we have considered optimization only with respect to virtual (unoccupied) space, having determined that this can be accomplished via diagonalization of the appropriate \(F_i\) or \(\tilde{F}_i\) matrices (OCBSE). We must now turn our attention to optimization with respect to occupied space, that is, we must deal with the problem of optimizing the orbitals with respect to one another.

5.2.1. Independent Pairwise Optimization

For simplicity we will first allow for mixing only between two orbitals at a time, so that

\[
\phi_i = (\phi_i^0 + \lambda_i \phi_i^0) / (1 + \lambda_i^2)
\]

\[
\phi_i = (\phi_i^0 + \lambda_i \phi_i^0) / (1 + \lambda_i^2)
\]

However, from (90) we have

\[
0 = \langle \phi_i^0 | \Delta \phi_i \rangle + \langle \phi_i^0 | \Delta \phi_i \rangle + \langle \Delta \phi_i | \Delta \phi_i \rangle
\]

\[
= \lambda_{ii} + \lambda_{ij}
\]

Thus, preservation of orbital orthogonality requires that

\[
\lambda_{ii} = - \lambda_{ij}
\]
so that

\[ \Delta \phi_i = \lambda_{ij} \phi_{ij}^0 \]
\[ \Delta \phi_j = -\lambda_{ij} \phi_{ij}^0 \]

(111)

Therefore, from (94a) we obtain

\[ \Delta \varepsilon_{ij}(1 + \lambda_{ij}^2) = 2\lambda_{ij} \langle i_0 | (F_{i_0} - F_{j_0}) | j_0 \rangle \]
\[ + \lambda_{ij}^2 \left[ \langle j_0 | (F_{i_0} - F_{j_0}) | j_0 \rangle - \langle i_0 | (F_{j_0} - F_{i_0}) | i_0 \rangle + \gamma_{ij} \right] \]

(112)

where

\[ \lambda_{ij} \gamma_{ij} = \langle i_0 | (\Delta F_{i_0} - \Delta F_{j_0}) | j_0 \rangle \]

(This makes \( \gamma_{ij} \) independent of \( \lambda_{ij} \).)

If we can take \( F_{i_0} = F_{j_0} \), then (112) reduces to

\[ \Delta \varepsilon_{ij} = 0 \]

which correctly reflects the fact that the energy is invariant with respect to mixing between orbitals in the same shell. Therefore, in cases where all orbitals belong to the same shell (e.g., an HF closed-shell singlet wave function), the iterative cycle consists solely of OCBSE optimization.

Assuming that orbitals \( i \) and \( j \) are in different shells, \( \lambda_{ij} \) is obtained by requiring that \( \Delta \varepsilon_{ij} \) be stationary with respect to \( \lambda_{ij} \),

\[ \frac{d(\Delta \varepsilon_{ij})}{d\lambda_{ij}} = 0 \]

to give

\[ 0 = (1 - \lambda_{ij}^2) \langle i_0 | F_{i_0} - F_{j_0} | j_0 \rangle - \lambda_{ij} \left[ \langle i_0 | F_{i_0} - F_{j_0} | i_0 \rangle - \langle j_0 | F_{i_0} - F_{j_0} | j_0 \rangle + \gamma_{ij} \right] \]

(113)

where from (91) and (92)

\[ \gamma_{ij} = 2(a_{ij} + a_{ji} - 2a_{ij})K_{ijij} + (b_{ij} + b_{ji} - 2b_{ij})(J_{ijij} + K_{ijij}) \]

(114)

In keeping with variational condition (84) if

\[ \langle i_0 | (F_{i_0} - F_{j_0}) | j_0 \rangle = 0 \]

then \( \lambda_{ij} = 0 \), since these orbitals are already optimal (self-consistent) with respect to each other. Assuming this is not the case, solving (113) for \( \lambda_{ij} \) gives

\[ \lambda_{ij} = \frac{-1}{2\lambda_{ij}^0} \left[ \left( \frac{1}{2\lambda_{ij}^0} \right)^2 + 1 \right]^{1/2} \]

(115a)
where
\[
\lambda_{ii}^0 = \frac{\langle i_0|F_{i0} - F_{i0}|j_0 \rangle}{\langle i_0|F_{i0} - F_{w0}|i_0 \rangle - \langle j_0|F_{j0} - F_{w0}|j_0 \rangle + \gamma_{ii}}
\] (115b)

Usually we want the solution leading to the lowest energy and therefore we choose the root which minimizes \(\Delta \varepsilon_{ii}\). However, under certain circumstances, such as in excited state calculations, a "least change" root selection criterion may be more appropriate.

Ignoring the \(\lambda_{ii}^2\) in (113) [which is equivalent to neglecting the renormalization term \((1 + \lambda_{ii}^2)\) in (112)] leads to
\[
\lambda_{ii} = \lambda_{ii}^0
\] (115c)

as the solution. This approximation can only be valid for \(\lambda_{ii}^2 \ll 1\). However, it is important to realize that (115c) may well result in \(\lambda_{ii}^2 > 1\). On the other hand, the "least change" root of (115a) [which reduces to (115c) in the limit of \(\lambda_{ii}^2 \ll 1\)] will supply a \(\lambda_{ii}^2 < 1\) result even when \((\lambda_{ii}^0)^2 > 1\). Moreover, as we shall see (112) can be sufficiently well behaved over a wide enough range for the \(\lambda_{ii}^2 > 1\) root of (115a) to be meaningful as well. Therefore, restriction of (115a) to the approximation (115c) is inappropriate since it robs us of important control over the iterative process and unnecessarily restricts the radius of convergence.

Being derived from (94), Eq. (113) is only an approximation to the exact equation for \(\lambda_{ii}\). A rigorous derivation of the expression for the change in energy under orthogonal mixing between orbitals \(i\) and \(j\) leads to the following (quartic) variational equation for \(\lambda_{ii}\),
\[
0 = \langle i_0|F_{i0} - F_{j0}|i_0 \rangle + \lambda_{ii}[\langle i_0|F_{i0} - F_{w0}|i_0 \rangle - \langle j_0|F_{j0} - F_{i0}|j_0 \rangle + \gamma_{ii}] \\
+ 3\lambda_{ii}^2\langle i_0|J_{i0} - J_{j0}|j_0 \rangle \alpha_{ii} \\
+ \lambda_{ii}^4[\langle i_0|F_{i0} - F_{w0}|i_0 \rangle - \langle j_0|F_{j0} - F_{w0}|j_0 \rangle + \gamma_{ii} + \alpha_{ii}(J_{i0} + J_{j0} - 2J_{i0} - 4K_{i00})] \\
- \lambda_{ii}^4[\langle i_0|F_{i0} - F_{w0}|i_0 \rangle + \alpha_{ii}(i_0|J_{i0} - J_{j0}|j_0 \rangle]
\] (115d)

where
\[
\alpha_{ii} = a_{ij} + b_{ij} + a_{ji} + b_{ji} - 2a_{ii} - 2b_{ii}
\]

Since we are only interested in the real roots for this equation, if the terms
\[
\alpha_{ii}\langle i_0|(J_{i0} - J_{j0})|j_0 \rangle
\] (115e)

and
\[
\alpha_{ii}(J_{i00} + J_{i00} - 2J_{i00} - 4K_{i00})
\]

were to vanish, then (115d) would reduce to (113). In several important instances this is in fact the case. For example, if one orbital is a simple singly
occupied orbital \((f_i = \frac{1}{2}, a_i + b_i = 0)\) and the other is a closed-shell or correlated pair orbital, then (113) is quite rigorous, since \(a_i = 0\). Even when these terms (115e) do not vanish, we generally find that the solutions to (115a) are sufficiently close to the real roots of (115d) so that the considerable added computational effort required to solve (115d) exactly is usually not justified. Bear in mind that we are considering here only the rotation* of two occupied orbitals and hence ignore the changes in the optimum \(\lambda_{ij}\) due to simultaneous rotations of other orbitals.

If the Fock correction term \(\gamma_{ij}\) were to vanish, then (115a) would be equivalent to diagonalizing the matrix

\[
(F_{ao} - F_{jo})
\]

over the space \(\{\phi_i^0, \phi_j^0\}\). Unfortunately this term can be quite important and should be retained since its neglect can reduce convergence and may even lead to oscillatory or divergent results.

Ideally, a pairwise optimization scheme would involve intermediate orbital redefinition. That is, once \(\lambda_{ij}\) has been determined orbitals \(i\) and \(j\) would be redefined before proceeding to the next pair. However, the computational ramifications of such a procedure are prohibitive due to the necessity of reprocessing the two-electron integrals. Therefore, we must use the same set of initial orbitals to calculate each mixing coefficient. Once this is done, the new orbitals are constructed from the results. There are many plausible ways of doing this. Of the various possibilities, we generally favor a sequential redefinition procedure based upon increasing coefficient magnitude.

Since each mixing coefficient is determined in a completely independent manner, it is not surprising to find that this method tends toward somewhat oscillatory behavior. However, this very simple procedure generally converges successfully.

### 5.2.2. Simultaneous Orbital Optimization

Pairwise orbital mixing, no matter how rigorously applied, represents maximum decoupling between all possible orbital changes within the space of \(\{\phi_i^0\}\). That is, a pairwise approach allows us to deal with only one degree of freedom at a time. Therefore, even though each pair mixing coefficient may be determined in a quadratically convergent manner, such a procedure can only be expected to lead to overall linear convergence. Furthermore, since we have variational control over only one variable at a time, we can expect to converge only to solutions directly accessible through individual pairwise changes. To overcome these limitations, it is necessary to allow orbital rotations to occur.

---

*We will often use the vector space analogy of a Hilbert space, referring to a mixing of orbitals as a rotation.*
simultaneously so that coupling between them can be taken into account. That is, we must consider the problem of optimizing all orbitals with respect to one another at the same time.

We will first reexpress the orbital expansions in terms of a set of linearly independent expansion coefficients. Ignoring (for ease of notation) same-shell invariances and possible orbital symmetry partitions, each orbital can be written in unnormalized form as

$$\phi_\nu = \sigma_0^\nu + \sum_{i=1}^n d_{i\nu} \phi_i^0$$

where we have ignored normalization since it drops out of the orthogonality conditions. However, preservation of orbital orthogonality requires that

$$0 = \langle \phi_i | \phi_j \rangle = d_{ij} + d_{ji} + \sum_{k \neq i, j} d_{ki} d_{kj} , \quad i \neq j$$

Hence, owing to the first-order term \((d_{ij} + d_{ji})\) appearing in (117), the variables \(\{d_{ij}\}\) do not constitute a linearly independent set. However, since first-order dependencies can be resolved by requiring that

$$d_{ij} = -d_{ji} \quad i \neq j$$

we are naturally led to an expression of the form

$$\phi'_\nu = \phi''_\nu + \sum_{i=1}^n \zeta_{i\nu} \lambda_{(ij)} \phi_i^0$$

where

$$\zeta_{ij} = -\zeta_{ji} = 1 \quad \text{for } i > j$$

and \((ij)\) is a pair index ranging from 1 to \(n(n-1)/2\),

$$(ij) = \left[ \max(i, j) - 1 \right] \left[ \max(i, j) - 2 \right]/2 + \min(i, j)$$

While (118) does provide us with an expansion in terms of the linearly independent variables \(\{\lambda_{(ij)}\}\), orthogonality is preserved only through first order. However, higher-order nonorthogonalities can easily be removed by orthogonalizing \(\{\phi_i''\}\) using the Gram–Schmidt procedure,

$$\phi_\nu = \phi''_\nu - \sum_{i < \nu} \frac{\langle \phi_i'' | \phi_i'' \rangle}{\langle \phi_i'' | \phi_i'' \rangle} \phi_i$$

Since explicit expansion of (119) through second order gives

$$\phi_\nu = \phi_\nu^0 + \sum_{i=1}^n \zeta_i \lambda_{(ij)} \phi_i^0 + \sum_{i=1}^n \sum_{j=1}^n \zeta_{ij} \lambda_{(ij)} \lambda_{(ij)} \phi_i^0 + O(\lambda^3)$$

(120)
we obtain the sought-after correction expansion,

\[ \Delta \Phi = \sum \xi_{i} \lambda_{(i)} \Phi_{i}^{0} + \sum \sum \xi_{i} \xi_{i} \lambda_{(i)} \lambda_{(i)} \Phi_{i}^{0} + O(\lambda^{2}) \]  

(121)

Having derived a correction expansion in terms of independent variables explicitly satisfying orthogonality condition (95) through second order, we can now apply (94) to obtain an optimization equation for \( \{ \lambda_{(i)} \} \). Substituting (121) into the various terms of (94a) and expanding each through second order in \( \{ \lambda_{(i)} \} \), we obtain

\[ \Delta \varepsilon = \left[ 1 + \sum \lambda_{(i)}^{2} \right]^{-1} \left\{ 2 \sum \xi_{i} \lambda_{(i)} \langle i_{0} | F_{0i} | j_{0} \rangle 
+ \sum \lambda_{(i)}^{2} \left[ \langle j_{0} | F_{0i} | j_{0} \rangle - \langle i_{0} | F_{0i} | i_{0} \rangle \right] 
+ 2 \sum \sum \xi_{i} \xi_{i} \lambda_{(i)} \lambda_{(i)} \langle i_{0} | F_{0i} | j_{0} \rangle 
+ \sum \sum \xi_{i} \xi_{i} \lambda_{(i)} \lambda_{(i)} \langle j_{0} | F_{0j} | k_{0} \rangle 
+ \sum \sum \sum \xi_{i} \xi_{i} \xi_{i} \lambda_{(i)} \lambda_{(i)} \lambda_{(i)} \langle j_{0} | \Delta F_{0i} | k_{0} \rangle \langle i_{0} | \Delta F_{0i} | j_{0} \rangle \right\} \]  

(122)

Unfortunately, despite the simplifying approximations already made the variational determination of \( \{ \lambda_{(i)} \} \) using (122) is still too complicated and we must make further approximations. The major difficulty with this equation is the disparity between the various orbital renormalization terms,

\[ \left[ 1 + \sum \lambda_{(i)}^{2} \right]^{-1} \]  

(123a)

Since we have already neglected many third-order and higher terms in arriving at (122), we could of course simply choose to ignore these terms and still retain second-order validity. However, by completely neglecting all reciprocal dependence upon \( \{ \lambda_{(i)} \} \) we would be forcing ourselves into a solution over which we could have no control and which would be valid only for a \( \lambda_{0}^{2} \ll 1 \) result. [See the discussion of (115a) and (115c).] This is a rather drastic step and one which we prefer to avoid. An alternate approach to this problem is to consider a formulation which, while making the necessary approximations to these renormalization terms, still preserves a fundamental reciprocal dependence with respect to \( \{ \lambda_{(i)} \} \). This can be accomplished by replacing these terms (123a) by the single term

\[ \left[ 1 + \sum \lambda_{(k)}^{2} \right]^{-1} \]  

(123b)
That is, instead of simply setting each of these denominators to a common fixed value of unity, we replace them with a common term exhibiting the correct functional dependence with respect to each individual mixing coefficient. In this way, we introduce additional errors only with respect to third-order (and higher) coupling between different mixing coefficients.

After making this common denominator approximation, evaluating Fock corrections, and collecting terms, Eq. (122) becomes

\[
\Delta \varepsilon [1 + \sum_{i>j} \lambda_{ij}^2] = \sum_{i>j} \{2\lambda_{ij} (\langle i_0 | F_{i_0} - F_{i_0} | j_0 \rangle) \\
+ \lambda_{ij}^2 [\langle i_0 | F_{i_0} - F_{i_0} | j_0 \rangle - \langle j_0 | F_{i_0} - F_{i_0} | j_0 \rangle + \gamma_{ij}^ll]
\]

\[
+ 2 \sum_{i,j,k \neq \ell} \{\lambda_{ij} \lambda_{jk} [\langle j_0 | F_{i_0} - F_{i_0} | k_0 \rangle + \gamma_{jk}^ll] \\
- \lambda_{ij} \lambda_{jk} [\langle i_0 | F_{i_0} - F_{i_0} | k_0 \rangle + \gamma_{jk}^ll] \\
+ \lambda_{jk} \lambda_{ik} [\langle i_0 | F_{k_0} - F_{k_0} | j_0 \rangle + \gamma_{ik}^kk] \}
\]

\[
+ 2 \sum_{i,j,k,l} \lambda_{ij} \lambda_{kl} \gamma_{kl}^ll
\]  

(124)

where

\[
\gamma_{kl}^ll = 2[a_{ik} + a_{il} - a_{il} - a_{ik}](i_0|k_0|l_0) + \gamma_{ik}^ll + \gamma_{ij}^ll
\]

(125)

If we define the symmetric matrix \( B \) by

\[
B_{0,0} = 0
\]

\[
B_{0,(ij)} = \langle i_0 | (F_{i_0} - F_{i_0}) | j_0 \rangle, \quad i > j
\]

\[
B_{(ij),(ij)} = [\langle i_0 | (F_{i_0} - F_{i_0}) | i_0 \rangle - \langle j_0 | F_{i_0} - F_{i_0} | j_0 \rangle + \gamma_{ij}^ll], \quad i > j
\]

\[
B_{(ij),(ik)} = \xi_{ij} \xi_{ik} [\langle j_0 | (F_{i_0} - F_{max_0}) | k_0 \rangle + \gamma_{ik}^ll], \quad i \neq j \neq k
\]

\[
B_{(ij),(kl)} = \gamma_{kl}^ll, \quad i > j, k > l, i \neq j \neq k \neq l
\]

and let

\[
C_{ij} = \lambda_{ij} C_0
\]

then (124) can be written as

\[
\Delta \varepsilon \sum_{i=0}^{m} C_i^2 = \sum_{i,j=0}^{m} C_i C_j B_{ij}
\]

(127a)
where \( nn = n(n-1)/2 \) (and where, for example, \( i \neq j \neq k \) is understood as \( i,j,k \), all different). Thus, the rotation expansion coefficients \( \{ C_i \} \), and hence the corrected orbitals \( \{ \phi_i \} \) for the current iteration, can be obtained by simply diagonalizing the \( B \) matrix and selecting the appropriate root,

\[
\sum_i B_{ij} C_i = C_i \Delta \varepsilon
\]

Equation (127b) allows for simultaneous orbital rearrangement within the space of \( \{ \phi_i^0 \} \) including the effects of each rotation \( C_{ij} \) due to simultaneous application of all other rotations \( C_{ik} \). Since (127b) is valid through second order for all mixing coefficients \( \{ \lambda_{ij} \} \) and since this expression couples the changes in all orbitals to one another, it is capable of leading to a controlled rapidly convergent solution. The coupling between individual mixing coefficients included here can be of crucial importance in permitting the essential simultaneous readjustments necessary to obtain the desired solution (some solutions are not readily accessible from a given set of trial functions \( \{ \phi_i^0 \} \) through individual pairwise considerations alone).

5.2.3. Individual Pairwise Optimization

Mixing all orbitals with respect to each other simultaneously is clearly the most unrestricted and desirable approach to optimizing orbitals within the minimal basis space of \( \{ \phi_i^0 \} \). However, even for the relatively simple approximation to this problem presented above, the computational effort involved in constructing (and diagonalizing) the \( B \) matrix required to properly couple the various mixing coefficients is prohibitive for general cases. Therefore, we usually assume that our starting guesses are reasonable enough so that convergence to a desired solution can be achieved through pairwise considerations alone. That is, while being fully aware of the limitations involved, we assume that coupling between individual pairwise changes can be ignored. Our problem thus reduces to finding the most efficient way of accomplishing this much simpler task.

We have already discussed the simplest possible approach to this problem in Section 5.2.1., whereby each mixing coefficient is obtained from \( \{ \phi_i^0 \} \) in a completely independent fashion using (115a) and an appropriate root selection criterion. While variationally determining \( \{ \lambda_{ij} \} \) in this manner usually does lead to successful results, this method has a tendency to exhibit somewhat oscillatory behavior and can at times converge relatively slowly. An alternate prescription that generally leads to superior results can be obtained from (127) by neglecting those elements of the \( B \) matrix that serve only to couple different mixing coefficients. That is, of the \( B_{ij} \) with \( i,j \neq 0 \) we include only the diagonal element. Within this approximation, our optimization equation reduces to

\[
\Delta \varepsilon [1 + \sum_{i,j} \lambda_{ij}^2] = \sum_{i,j} [2 \lambda_{ij} B_{0,ij} + \lambda_{ij}^2 B_{ii,ij}] \]

(128)
Since we need only evaluate the matrix elements \( \{ B_{0,(ij)}; B_{(ij),(ii)} \} \) (no other off-diagonal elements), this approximation greatly simplifies matters relative to (127). In fact, since (115b) can be written as
\[
\lambda_{ij}^0 = -\frac{B_{0,(ij)}}{B_{(ii),(ii)}}
\]
we see that (128) requires no more information than is needed for an independent pairwise mixing procedure. The resulting matrix \( B \) is quite sparse (two nonzero elements per row) and hence this matrix can be diagonalized rapidly.\(^{(18)}\)

While (128) does not involve any direct coupling between different mixing coefficients (and is therefore still only valid through first order, overall), in solving this equation for \( \{ \lambda_{(ij)} \} \) through diagonalization of the resultant approximate \( B \) matrix, each coefficient is nonetheless linked to all others in such a way that each is determined as "weighted" with respect to its relative contribution to \( \Delta \varepsilon \). Since this contribution is at least a gross reflection of the relative energy contribution for each pairwise change, this weighting is a desirable property. In evaluating each coefficient independently as in (115a), however, even this most minimal correlation is not possible. As a consequence, it is not surprising that an independent evaluation scheme (115a) tends to overestimate total orbital changes* and can therefore lead to oscillatory behavior. In solving (128) instead, those changes being of greatest importance are weighted more heavily than all others, thereby minimizing this effect in a quite natural manner.

### 5.2.4. Summary of Orbital Mixing Considerations

Inclusion of all coupling terms in the energy expression through second order leads to the variational Eqs. (127) which include the effect on each rotation \( C_{(ij)} \) of the simultaneous rotations of all other orbital pairs \( \{ C_{(ki)} \} \). The resulting Eqs. (127) is thus the proper one for iteratively solving for the optimum orbitals. Indeed we could also include in (127) the rotations between occupied and virtual orbitals and hence obtain quadratic convergence. Unfortunately at present (127) seems too costly, despite the rapid convergence to be expected.

At the other extreme we considered in (115a) the rotation of one pair of orbitals at a time. To sequentially solve rigorously for each rotation is impractical at the present time since the two-electron integrals must be reprocessed after each rotation. Thus when using (115a) we generally take a set of starting orbitals and solve for all \( n(n-1)/2 \) rotation coefficients \( \lambda_{(ij)} \), each of which assumes the same set of starting coefficients. These rotations are not all consistent with each other and some procedure of sequentially applying them

\*This effect can sometimes be used to our advantage in providing us with fortuitous extrapolatory results that can lead to an increased rate of convergence.
must be selected. This approach has been generally successful but is often to blame for the slow convergence that sometimes occurs.

Our compromise between these two extremes is (128), which determines all the rotations simultaneously but utilizes only quantities already required for (115a). In addition, (128) leads to no special storage or matrix diagonalization difficulties. Although determining the rotations simultaneously, this approach neglects the special effects upon the energy of simultaneous rotation of the orbitals. As a result, although rapidly convergent it cannot be expected to converge quadratically.

Because of its good overall convergence capabilities and computational simplicity, the orbital mixing optimization approximation (128) developed here has been found to be of excellent general utility. This procedure is also quite compatible with the OCBSE optimization method described previously in that both steps have similar computational requirements and handle their respective tasks to within almost the same degree of overall validity. Therefore, in the union of these two complementary steps we obtain a practical and yet highly efficient iterative prescription for obtaining SCF solutions for the orbitals in wave functions of the type we have been dealing with.

### 5.3. Correlated Pair Coefficient Optimization

For wave functions whose energy expressions involve fixed energy parameters \( \{f_n, a_{ij}, b_{ij}\} \), self-consistency is achieved solely through orbital optimization, as described in Sections 5.1 and 5.2. Such is the case for HF wave functions. However, for wave functions containing correlated pairs, such as the GVB wave function in (58), these parameters are functions of the variable pair coefficients \( \{\sigma_{ij}\} \). Therefore, in order to obtain fully self-consistent solutions for such wave functions, these coefficients must also be variationally determined.

### 5.3.1. The General Case

This task is most easily accomplished by adding yet a third step to the iterative cycle. In this step we optimize \( \{\sigma_{ij}\} \) while keeping the orbitals fixed. To see how this is done let us focus attention on a single such pair. Abstracting from the total energy only those terms that depend on the \( n \)th correlated pair, we obtain

\[
E_v = \sum_{i=1}^{n_v} \left[ 2f_{iv} h_{iv,iv} + (a_{iv,iv} + b_{iv,iv}) J_{iv,iv} \right] + \sum_{j \neq 1} (a_{iv,jv} J_{iv,jv} + b_{iv,jv} K_{iv,jv}) \right] + \sum_{i \neq j} (a_{iv,jv} J_{iv,jv} + b_{iv,jv} K_{iv,jv}) \right]
\]

(129)
where we have allowed \( n_v \) natural orbitals in the correlated pair (\( n_v = 2 \) for GVB). Using (59) to write (129) in terms of \( \{ \sigma_{i,v} \} \) leads to

\[
\frac{1}{2}E_v = \sum_{i,l} \sigma_{i,v} Y_{i,l}^v
\]

(130a)

where

\[
Y_{i,l}^v = h_{i,v} + \frac{1}{2}J_{i,v} + \sum_{j \neq i, 1v, 2v} f_i(2J_{j,v} - K_{j,v})
\]

\[
Y_{i,v} = \frac{1}{2}K_{i,v} \quad \text{for } i \neq j
\]

and

\[
\sum_i \sigma_{i,v}^2 = 1
\]

Therefore, \( \{ \sigma_{i,v} \} \) can be obtained by simply diagonalizing the \( Y^v \) matrix,

\[
Y^v \sigma_v = \varepsilon_v \sigma_v
\]

(130b)

and normally choosing the lowest energy root.

Since the diagonal elements of the \( Y^v \) matrix depend upon \( \{ f_i/j \neq 1v, 2v, \ldots \} \), the \( \{ \sigma_{i,v} \} \) must be iteratively determined for wave functions involving more than one correlated pair. This can be most rapidly accomplished through a sequential procedure involving intermediate redefinition of these matrices.

5.3.2. Simplifications for GVB Pairs

If we are dealing with a GVB pair, then there are only two orbitals involved and the pair coefficients \( \sigma_{1,v} \) and \( \sigma_{2,v} \) can be determined in a very simple manner since the pair energy contribution is simply

\[
2E_v = (\sigma_{1,v}^2 Y_{1,1}^v + \sigma_{2,v}^2 Y_{2,2}^v + \sigma_{1,v} \sigma_{2,v} K_{1v,2v}) / (\sigma_{1,v}^2 + \sigma_{2,v}^2)
\]

(131)

[Note that the third term of (50) differs from that of (131) because we had built a minus sign into (49).] Therefore, letting

\[
\alpha_v = \sigma_{2,v} / \sigma_{1,v}
\]

and deleting from (131) the constant \( Y_{1,1}^v \), we obtain

\[
\mathcal{E}_v = (\alpha_v^2 Y_v - \alpha_v K_{1v,2v}) / (1 + \alpha_v^2)
\]

(132)

where

\[
\mathcal{E}_v = 2E_v - Y_{1,1}^v
\]

and

\[
Y_v = Y_{2,2}^v - Y_{1,1}^v
\]

Requiring the energy to be stationary with respect to \( \alpha_v \), \( d\mathcal{E}_v/da_v = 0 \), and assuming that \( K_{1v,2v} \neq 0 \) (\( K_{1v,2v} = 0 \) results in collapse of the pair), leads to a quadratic equation having the solutions

\[
\alpha_v = \frac{Y_v}{K_{1v,2v}} \pm \left[ 1 + \left( \frac{Y_v}{K_{1v,2v}} \right)^2 \right]^{1/2}
\]

(133)
However, since it is always the case that $K_{1v,2v} = 0$, the root leading to the most negative energy must correspond to $\alpha_v \leq 0$. Therefore, we obtain the simple final result

$$\sigma_{1v} = 1/(1 + \alpha_v^2)^{1/2}$$

$$\sigma_{2v} = \alpha_v/(1 + \alpha_v^2)^{1/2}$$

(134)

where

$$\alpha_v = \frac{Y_v}{K_{1v,2v}} \left[ 1 + \left( \frac{Y_v}{K_{1v,2v}} \right)^2 \right]^{1/2}$$

(135)

### 5.4. The Iterative Sequence

The computer programs implementing the variational methods described in the previous sections are continually evolving and hence likely to change significantly over the next few years. Even so, it is appropriate to outline the steps involved in the current programs (referred to as GVBTWO*) in order to provide the reader with a clearer picture of how to put together the various steps.

#### 5.4.1. Initialization

*Energy Coefficients. The energy coefficients $(f_n, a_i, b_i)$ for the common cases such as described in Section 3 are built into the program so that only the spin state and identity of open-shell orbitals need be supplied. For correlated-pair wave functions, one provides the number of pairs, the number of orbitals in each pair and initial guesses for the pair coefficients $\{\sigma_{pi}\}$ from which the program calculates the energy coefficients.

*Transformation Matrix. Each occupied orbital $\phi_i$ is initially specified in terms of the atomic basis (AO) employed $\{\chi_i\}$ over which the one- and two-electron integrals have been evaluated,

$$\phi_i = \sum_j \chi_j T_{ji}$$

The program constructs orthonormal virtual orbitals (if they are not provided). The full set of orthonormal orbitals (occupied and virtual) is referred to as the MO basis. This basis is partitioned into symmetry types since the program is designed to take full advantage of the fact that orbital symmetries will be

*Developed by F. W. Bobrowicz, W. R. Wadg, and W. A. Goddard III (Spring 1973) (a more detailed description is given in Ref. 4). An earlier version, referred to as GVBOONE, is outlined in Ref. 2.
preserved during the iterations. Each occupied orbital is identified in terms of
the shell or correlated pair to which it belongs.

c. Initial J and K Matrices. For each open shell, the initial Coulomb and
exchange matrices are constructed over the AO basis*

\[ J_{\mu\nu} = \langle \chi_\mu | J | \chi_\nu \rangle \]
\[ K_{\mu\nu} = \langle \chi_\mu | K | \chi_\nu \rangle \]

Likewise an initial \( h^c \) matrix is constructed over AOs

\[ h^c_{\mu\nu} = \langle \chi_\mu | h | \chi_\nu \rangle + \langle \chi_\mu | \sum_p^{\text{closed}} (2J_p - K_p) | \chi_\nu \rangle \]

In these matrices only those \( \mu\nu \) elements are evaluated that are required
from the subsequent construction of the symmetry block matrices (step a
below). That is, only if both \( \mu \) and \( \nu \) can appear in the expansion of two orbitals
of the same symmetry.

5.4.2. The Iterative Process

a. J and K Matrices over MOs. The \( h^c \) and \( \{J_i, K_i\} \) matrices over AOs are
transformed to symmetry blocks of occupied orbitals (i.e., only those elements
are calculated in which both components belong to the same symmetry).

b. New Correlated-Pair Coefficients. For each pair \( i \), the \( Y \) matrices
defined in (130a) are constructed and the elements required for updating (vide
infra) these matrices are also stored conveniently. Each matrix is diagonalized
as in (130b) or solved as in (135) to obtain new pair coefficients \( \{\sigma_{ii}\} \). The \( Y \)
matrices for subsequent pairs are updated to reflect the new \( \{\sigma_{ii}\} \) for pair \( i \). This
process is continued until it converges. Based on the final new pair coefficients
\( \{\sigma_{ii}\} \), new energy coefficients \( \{f_i, a_i, b_i\} \) are defined for subsequent steps.

c. Orbital Mixing. Using the matrices over MOs from step a, the sparse \( B \)
matrix of (128) is constructed. Elements connecting orbitals of the different
symmetry or connecting orbitals of the same shell are not included. The matrix
\( B \) is diagonalized selecting either the lowest energy root or else the root
representing the least change. From the result, new occupied orbitals are
defined; however, the \( h^c \) and \( \{J_i, K_i\} \) over AOs are not modified.

*Rather than the Coulomb and exchange matrices discussed in the text, the program actually deals
with the modified quantities

\[ J = \frac{1}{2}(J - \frac{1}{2}K_i) \]
\[ X_i = \frac{1}{4}K_i \]

because these matrices can be constructed much more rapidly using a suitably preprocessed list of
two-electron integrals analogous to the approach developed by Raffenetti.\(^{(19)}\)
d. OCBSE. For each shell we go through the following sequence. (i) Construct the F matrix (over AOs) from the \( h^c \), \( \{J_n, K_i\} \) matrices; (ii) transform F to MOs including only those symmetry blocks that contain orbitals of this shell; (iii) diagonalize each symmetry block and select roots either on the basis of eigenvalues or else of least change; (iv) define new virtuals for each symmetry affected by (iii); (v) reconstruct new matrices over AOs for this shell (\( h^c \) if it is the closed shell or else the appropriate \( \{J_n, K_i\} \) if it is an open shell); (vi) proceed to the next shell.

e. If not converged, go back to step a.

6. Summary

We have considered herein general procedures of optimizing, self-consistently, the orbitals of Hartree–Fock and generalized valence bond wave functions. In particular, we considered all wave functions with energies that can be expressed as (6),

\[
E = 2 \sum_i f_i h_{ii} + \sum_{i,j} (a_{ij} J_{ij} + b_{ij} K_{ij})
\]

although this technique and formalism are valid for more general cases.

Requiring that the energy be stationary with respect to all orbital variations leads to the general variational condition in (68),

\[
\sum_i \langle \delta \phi_i | F_i | \phi_i \rangle = 0
\]

(68)

where \( F_i \) is the generalized Fock operator,

\[
F_i = f_i h + \sum_j (a_{ij} J_{ij} + b_{ij} K_{ij})
\]

(66)

Considering variations in which only orbital \( \phi_v \) changes with all other \( n - 1 \) orbitals kept fixed, (68) leads to

\[
\langle \delta \phi_v | F_v | \phi_v \rangle = 0
\]

(76)

Ignoring the orthogonality conditions (76), for a complete set of functions, leads to the pseudoeigenvalue equation

\[
F_v \phi_v = \epsilon_v \phi_v
\]

(136)

the more common way of expressing this condition. Although often stated as the variational condition, (76) is a necessary but not sufficient condition for optimum orbitals. Allowing occupied orbitals to mix with each other leads to
the condition
\[ \langle \phi_\nu | (F_\nu - F_\mu) | \phi_\mu \rangle = 0 \]  
(84)

for all occupied orbitals \( \mu \) and \( \nu \). Together (76) and (84) form necessary and sufficient conditions for a set of orbitals \( \{ \phi_i \} \) to yield a stationary energy.

Although the variational conditions (76) and (84) tell us whether a given set of orbitals are optimum (i.e., lead to stationary energy), they do not tell us what to do to improve the orbitals when they are not optimum. When using (136), a common approach is to solve
\[ F_{\nu \mu} \phi_\mu = \varepsilon_\nu \phi_\nu \]

where \( F \) is evaluated using the trial functions [this requires the second condition, (84)]. We believe that this is not a reliable approach and instead return to the total energy (the quantity we are trying to minimize), expand it in terms of simultaneous corrections in all orbitals,
\[ \phi_i = (\phi_i^0 + \Delta \phi_i)/(1 + (\Delta \phi_i|\Delta \phi_i))^{1/2} \]

keeping terms through second order leading to (94a) with (95) as the condition ensuring orbital orthonormalities. To obtain the optimum orbitals, we require that \( \partial E/\partial \Delta \phi_i = 0 \) for all variations consistent with orbital orthonormalities, leading to an equation analogous to (127),
\[ \sum_{(kl)} B_{(ij),(kl)} C_{(kl)} = C_{(ij)} \Delta \varepsilon \]  
(137)

where \( B \) is given by (126) and \( C_{(ij)} \) determines the mixing of orbitals \( \phi_i \) and \( \phi_j \). In addition, to one element for each of the \( n(n-1)/2 \) rotations, (137) contains an element \( C_0 \) corresponding to a renormalization term and the rotations are given by (126b),
\[ \lambda_{(ii)} = C_{(ii)}/C_0 \]  
(126b)

The final orbitals are given as
\[ \phi_\nu = (\phi_\nu^0 + \Delta \phi_\nu)/(1 + (\Delta \phi_\nu|\Delta \phi_\nu))^{1/2} \]

where
\[ \Delta \phi_\nu = \sum_{i \neq \nu} \zeta_{iv} \lambda_{(iv)} \phi_i^0 + \sum_{i < v} \sum_{j \neq i, \nu} \zeta_{ij} \xi_{jk} \lambda_{(ij)} \lambda_{(jk)} \phi_i^0 \]  
(121)

and
\[ \zeta_{\nu} = -\zeta_{\mu} = 1 \quad \text{for } i > j \]

Since (137) was derived from an energy expression correct through second order in all orbital corrections, then the errors in (137) are of order \( \Delta^2 \) and hence the orbital corrections \( C_{(kl)} \) will converge quadratically.
Unfortunately (137) seems inordinately costly to evaluate and to solve and hence we use a slightly more approximate approach. First we partition the problem into two parts

1. OCBSE: optimizing each orbital (sequentially) allowing it to mix only with the virtual orbitals.
2. Orbital mixing: optimizing the mixing of various occupied orbitals with each other.

The OCBSE leads to a simple eigenvalue equation (105b)

$$\tilde{F}_v C_v = C_v \varepsilon_v$$  \hspace{1cm} (105b)

for the new orbitals (in the space excluding other occupied orbitals).

Including all orbital mixing terms consistently leads to an equation (127), analogous to (137) above. This equation can be simplified to the form (128) at great saving in cost but with only slight penalties in convergence.

For wave functions of the GVB and separated-pair types in which the energy parameters in (6) depend on other expansion coefficients, these coefficients are solved for separately as in (130) or (135).

### 7. Comparison with Previous Work

As described in Section 4.2, a necessary and sufficient condition for a set of orbitals \( \{ \phi_i \} \) to be optimal is that the first-order change in the energy be zero for all changes in these orbitals consistent with those constraints (i.e., orbital orthonormality) imposed when deriving the energy expression. For real orbitals this results in the variational expression

$$\sum \langle \delta \phi_i | F_v | \phi_i \rangle = 0$$ \hspace{1cm} (68)

Until 1969, the usual approach to this problem was to allow only one orbital, say \( \phi_v \), to vary at a time keeping the remaining \( n - 1 \) orbitals fixed. This leads to

$$\langle \delta \phi_v | F_v | \phi_v \rangle = 0$$ \hspace{1cm} (76)

for all variations \( \delta \phi_v \) consistent with the constraints

$$\langle \delta \phi_v | \phi_i \rangle = 0 \quad \text{for all } \phi_i$$ \hspace{1cm} (78)

that is, all \( \delta \phi_v \) orthogonal to the occupied orbitals. If we substitute \( \delta \phi_v = \lambda \phi_i \) in (76) [that is, allow \( \delta \phi_v \) not satisfying (78)], the result is

$$\langle \delta \phi_v | F_v | \phi_v \rangle = \lambda \langle \phi_i | F_v | \phi_v \rangle = \varepsilon_{iv} \langle \delta \phi_v | \phi_i \rangle$$

where

$$\varepsilon_{iv} \equiv \langle \phi_i | F_v | \phi_v \rangle$$ \hspace{1cm} (138)
is referred to as a Lagrange multiplier. Thus (76) can be rewritten as

\[ \langle \delta \phi_{\nu}|F_{\nu}|\phi_{\nu}\rangle - \sum_{i} \epsilon_{i\nu} \langle \delta \phi_{\nu}|\phi_{i}\rangle = 0 \]  \hspace{1cm} (139)

which is valid for all \( \delta \phi_{\nu} \) no constraints. Visualizing \( \langle \delta \phi_{\nu}|F_{\nu}|\phi_{\nu}\rangle \) as the derivative of the energy in the direction \( \delta \phi_{\nu} \) we see that the Lagrange multiplier \( \epsilon_{i\nu} \) serves to annihilate that part of the energy derivative arising from varying \( \phi_{\nu} \) into the forbidden direction \( \phi_{i} \). In order for (139) to be satisfied for all \( \delta \phi_{\nu} \), the \( \{ \phi_{i}\} \) must satisfy

\[ F_{\nu} \phi_{\nu} = \sum_{i} \epsilon_{i\nu} \phi_{i} \]  \hspace{1cm} (140)

which can be taken as the fundamental variational equation.\(^{20} \) Nothing is changed here of course; (140) is equivalent to (76) with (78) and still only ensures that the energy is stationary upon varying only \( \phi_{\nu} \) in the space orthogonal to all occupied orbitals.

Substituting (138) into (140) and rearranging terms leads to the pseudo-eigenvalue equation of Birss and Fraga,\(^{21} \)

\[ R_{\nu} \phi_{\nu} = \epsilon_{\nu\nu} \phi_{\nu} \]  \hspace{1cm} (141a)

where

\[ R_{\nu} = F_{\nu} - \sum_{i=\nu} |\phi_{i}\rangle \langle \phi_{i}|F_{\nu} \]  \hspace{1cm} (141b)

"Integrating" (139) leads to a new functional,

\[ 2I_{\nu} = E - \epsilon_{\nu\nu} \langle \phi_{\nu}|\phi_{\nu}\rangle - 2 \sum_{i=\nu} \epsilon_{i\nu} \langle \phi_{i}|\phi_{\nu}\rangle \]  \hspace{1cm} (142)

where \( E \) is the energy expression of (6) and the factor \( \frac{1}{2} \) with the \( E \) arises from the definition of \( F_{\nu} \) in (65) and (66). Requiring that the first-order variation \( \delta I_{\nu}^{(1)} \) of (142) be zero for all \( \delta \phi_{\nu} \) leads directly to the correct variational equation (140). Thus, a shortcut to deriving the variational equations is to construct the functional (142) (where the \( \epsilon_{i\nu} \) are yet undetermined), apply the variational principle to obtain (140), and project (140) onto the functions \( \{ \phi_{i}\} \) to obtain (138). This sequence of steps is referred to as Lagrange's method of undetermined multipliers. However, it should be kept in mind that the fundamental variational condition is (76) with (78) or equivalently (140) with (138).

Consider now the variational equation for a second orbital \( \phi_{\mu} \). The equations corresponding to (76)-(78) and (138)-(140) lead to

\[ F_{\mu} \phi_{\mu} = \sum_{i} \epsilon_{i\mu} \phi_{i} \]  \hspace{1cm} (143)

where

\[ \epsilon_{i\mu} = \langle \phi_{i}|F_{\mu}|\phi_{\mu}\rangle \]  \hspace{1cm} (144)
which can be integrated to the functional

\[ 2I_\mu = E - \varepsilon_{\mu \mu} \langle \phi_\mu | \phi_\mu \rangle - 2 \sum_{j \neq \mu} \varepsilon_{\mu j} \langle \phi_j | \phi_\mu \rangle \]  

(145)

We may now consider the question of whether there might be some simple functional, say*

\[ 2I = E - \sum_{j,k} \varepsilon_{jk} \langle \phi_j | \phi_k \rangle \]  

(146)

that could be used to derive all \( n \) variational equations [such as (140) and (143)]. The answer, in general, is \textit{no}. For example, the variational condition from \( \delta \phi_\mu \) requires, letting \( j = \mu \) in (138),

\[ \varepsilon_{\mu \nu} = \langle \phi_\mu | F_\nu | \phi_\nu \rangle \]  

(147)

whereas the variational condition from \( \delta \phi_\mu \) requires [letting \( j = \nu \) in (144)]

\[ \varepsilon_{\nu \mu} = \langle \phi_\nu | F_\mu | \phi_\mu \rangle \]  

(148)

These quantities are not equal in general and hence the functional (146) does not lead to the correct variational equations, e.g., (140) and (143).

Unfortunately many workers have assumed that (146) is the proper starting point for applying the method of Lagrange multipliers to SCF problems. This has led to some confusion in the literature. Birss and Fraga\(^{(21)}\) noticed that the conditions (147) and (148) appear different but believing from (146) that these quantities must be equal, they assumed that

\[ \langle \phi_\mu | F_\nu | \phi_\nu \rangle = \langle \phi_\mu | F_\mu | \phi_\mu \rangle \]

or

\[ \langle \phi_\mu | (F_\nu - F_\mu) | \phi_\nu \rangle = 0 \]  

(149)

must automatically be satisfied for the self-consistent solutions of (76) and (78) [all \( \nu \)]. Das\(^{(22)}\) recognized that (149) may not be satisfied and suggested two modifications: (i) symmetrization of the operator \( R_\nu \) in (141b) and (ii) rotating each orbital pair during the iterative process so as to guarantee that (149) is satisfied. Levy\(^{(25)}\) attempted to incorporate this condition into the pseudoeigen-value problem itself by generalizing the operator \( R_\nu \) to have the form

\[ R_\nu^* = F_\nu - \sum_{j \neq \nu} \left[ \langle \phi_j | \phi_\nu \rangle G_{\mu \nu} + G_{\nu \mu} \langle \phi_\nu | \phi_j \rangle \right] \]  

(150a)

*Since \( \langle \phi_i | \phi_k \rangle = \langle \phi_k | \phi_i \rangle \), we have arbitrarily included the off-diagonal terms twice.
where

\[ G_{j\nu} = G_{\nu j} = \lambda F_{\nu} + (1 - \lambda) F_{j} \quad \text{for} \ j > \nu \]  

(150b)

and where \( \lambda \) is an arbitrary parameter. Hinze and Roothaan\(^{26}\) suggested averaging the \( \epsilon_{ij} \) from (147) and (148), leading to

\[ \epsilon_{ij} = \langle \phi_{i} | \frac{1}{2} (F_{i} + F_{j}) | \phi_{j} \rangle \]  

(151)

[i.e., \( \lambda = \frac{1}{2} \) in (150b)]. Finally Huzinaga\(^{26}\) suggested use of an even more general operator

\[ R_{j}^{H} = F_{\nu} - \sum_{i \neq \nu} [\langle \phi_{i} | \langle \phi_{j} | G_{j\nu} + G_{\nu j} | \phi_{j} \rangle | \phi_{j} \rangle] \]  

(152a)

where

\[ G_{j\nu} = \lambda_{j\nu} F_{j} + (1 - \lambda_{j\nu}) F_{\nu} \]  

(152b)

(all \( j \) and \( \nu \)) and where \( \lambda_{j\nu} \) is an arbitrary parameter.

It is important to keep in mind that the above formulations are all still based on the variational conditions arising from varying only one orbital at a time. Requiring that (149) also be satisfied was considered as an auxiliary condition which was felt to be necessary since otherwise the variational equations were not consistent with (146).

In 1969 Hunt et al.\(^{27}\)* pointed out that a self-consistent set of orbitals satisfying condition (76) with (78) need not satisfy (149). Thus, they pointed out that the usual functional (146) does not lead to the proper variational equations. However, they soon learned that solving self-consistently for the orbitals satisfying (76) with (78) leads to different solutions depending upon the trial functions.

These difficulties were finally resolved by Goddard et al.\(^{30}\) in 1969 who showed that although a necessary condition for a set of orbitals to be optimum is that (76) with (78) be satisfied for each orbital; these conditions do not constitute sufficient conditions. One must also consider variations in which each orbital is allowed to incorporate character of any other occupied orbital as well. As

---

*This is also the paper that proposed the OCBSE method as a simple approach to eliminating the need for Lagrange multipliers or the coupling operators with which they were often replaced. The basic idea in OCBSE of using the current set of occupied and virtual orbitals as the basis in terms of which the SCF equations are solved had been suggested earlier (1961) by Huzinaga\(^{28}\) (method II) but was apparently never applied. Probably this was because Huzinaga's approach involved carrying out a full transformation over the two-electron integrals to the new basis, a quite impractical procedure to carry out every iteration. Hunt et al.\(^{27}\) showed that one could evaluate the usual Coulomb and exchange operators over atomic orbitals (no integral transformation) just as if there were no constraints and then transform the final Fock matrices to the OCBSE basis, a quite practicable procedure. The OCBSE procedure was also rediscovered by Peters\(^{29}\) in 1972.
discussed in Section 4.2, because of the orthogonality constraint such a
variation must allow the other occupied orbital to simultaneously rotate out of
the way. Hence one is obligated to allow at least two orbitals to change
simultaneously and cannot expect to solve the problem through single-orbital
variations alone. Allowing all such variations of \( \phi \) and concurrent mandated
variations in all other orbitals, these authors obtained the equation

\[
R \phi = \epsilon \phi \tag{153a}
\]

where

\[
R = F - \sum_{j=\nu} |\phi_j\rangle \langle \phi_j| F
\tag{153b}
\]

[note well the presence of \( F_i \) in (153b) in place of the \( F \) of (141b)]. Projecting
(153a) onto \( \phi_k \) we obtain

\[
\langle \phi_k | R \phi \rangle = \langle \phi_k | F \phi \rangle - \langle \phi_k | F_k \phi \rangle = \langle \phi_k | F_k - F_\nu \phi \rangle = 0
\]

which is the condition derived in (84). Comparing with (149) we see that
optimizing the mixing of the occupied orbitals with each other does indeed lead
to equal off-diagonal Lagrange multipliers, (147) and (148), the condition
usually assumed.

Dahl et al.\(^{31}\) showed that starting with (146) and (149) leads to the
equations derived by Goddard et al.\(^{30}\) and disputed the latter authors' claim of
an additional necessary condition. The error in reasoning by Dahl et al. was
pointed out in the literature by Albat and Gruen\(^{32}\) and (153) is recognized\(^{33}\)
as the proper variational equation.

Although (153) are the proper variational conditions, satisfaction of which
leads to fully self-consistent orbitals, there are a number of ways of attempting
to iteratively solve for these orbitals. One such procedure was proposed by
Davidson\(^{34}\) who based his iterative procedure on a symmetrized form of
(153b). In so doing it was noted that some sort of damping was required in
order to attain smooth convergence. Wood and Veillard\(^{35}\) also implemented
this equation, using level shifting techniques to overcome convergence difficul-
ties.

In implementing the iterative solutions of pseudoeigenvalue equations
such as (141) or (153),

\[
R \phi = \epsilon \phi \tag{154}
\]

the usual approach had been to evaluate the operator \( R \) in terms of a set of
trial functions \( \{ \phi^0 \} \) and then to solve equations such as

\[
R_{\nu \nu} \phi = \epsilon \phi \tag{155}
\]
for the new orbitals; this process being repeated until it converges. In 1970 Hunt et al.\(^6\) pointed out the inconsistency of this approach, showing that in solving for the mixing coefficient for \(\Phi_v\) and \(\Phi_\mu\) one obtains two different results depending upon whether one starts with the equation for \(\Phi_v\):

\[
R_{\nu 0} \Phi_v = \varepsilon \Phi_v
\]

or the equation for \(\Phi_\mu\),

\[
R_{\mu 0} \Phi_\mu = \varepsilon_\mu \Phi_\mu
\]

Hunt et al.\(^6\) resolved this problem by formulating a set of equations based directly upon variation condition (84) to obtain a set of coupled equations,

\[
\sum_{(k,l)} B_{(i)(kl)} \Delta_{kl} = -X_{(i)}
\]

for the linearly independent orbital corrections \(\{\Delta_{kl}, k > l\}\). Since these equations neglected only terms to second order in \(\Delta\), they lead to quadratic convergence. Hunt et al.\(^6\) then simplified these equations by partitioning the problem into an OCBSE step and an orbital mixing step. For the orbital mixing step they considered independent pairwise rotations as in Section 5.2.1, using the linear Eq. (115c). This approach was the basis of the GVBONE program of Hunt et al.\(^3\)

Our current approach to deriving the variational equations differs from that of Hunt et al.\(^6\) in that we base all considerations upon the total energy expression itself, a starting point that has largely been ignored previously.\(^*\) The energy is expanded through second order in the orbital corrections (including orthogonality constraints) and the optimum corrections are obtained from the requirement that \(\frac{\partial E}{\partial \Delta \Phi_i} = 0\) for all \(i\). In addition, we have generalized the orbital mixing steps as indicated in Section 5.2.

**Acknowledgment**

This work was partially supported by a grant (MPS 74-05132) from the National Science Foundation.

\(^*\)Rossi\(^{36}\) in 1967 suggested solving for SCF orbitals by applying the \(n(n - 1)/2\) two-by-two rotations among the various orbitals. This formalism was based upon varying the total energy directly. Mukherjee\(^{37}\) attempted to develop this approach but unfortunately did not retain sufficient terms to satisfy the variational conditions!
References