

# Approximate second order method for orbital optimization of SCF and MCSCF wavefunctions

Galina Chaban, Michael W. Schmidt, Mark S. Gordon

Department of Chemistry, Iowa State University, Ames, IA 50011, USA

Received: 15 August 1996 / Accepted: 22 January 1997

**Abstract.** A quasi-Newton method involving a diagonal guess orbital hessian with iterative updates has been proposed recently by Almlöf for the optimization of closed shell self-consistent field (SCF) wavefunctions. The technique is extended in the present work to more general wavefunctions, ranging from open shell SCF through multiconfigurational SCF. A number of examples are presented to show that convergence for closed and open shell SCF rivals conventional direct inversion in the iterative subspace (DIIS). For multiconfigurational SCF wavefunctions, the method presented here requires more iterations than an exact second order program, but since each iteration is substantially faster, leads to a more efficient overall program.

**Key words:** Optimization of orbitals – Open shell SCF – GVB – MCSCF

## 1 Introduction

Most quantum chemistry calculations start from Hartree-Fock (HF) self-consistent field (SCF), generalized valence bond (GVB), or multiconfigurational self-consistent field (MCSCF) wavefunctions, which serve as starting points for obtaining more accurate wavefunctions that include electron correlation by means of configuration interaction, cluster expansions, or perturbation theory.

Optimization of molecular orbitals expanded in a basis of atomic orbitals is the common step that is required to obtain all of these wavefunctions. The standard scheme for SCF orbital optimization is based on an iterative procedure involving diagonalization of the Fock matrix [1] normally supplemented today by a method referred to as direct inversion in the iterative subspace (DIIS) [2–4]. An alternative approach with an almost equally long history [5–16] is a second order

method involving Newton-Raphson optimization. This method uses an exponential parameterization to obtain orbital rotation angles, and requires calculation of first and second derivatives of the energy with respect to these orbital coordinates (orbital gradient and orbital hessian). This method has not been very popular for optimization of HF wavefunctions, due to the high cost of computing the orbital hessian (which requires an integral transformation in the original formulations [8, 9]), as well as the large size of the hessian matrices which are difficult to store even for average size molecular systems. A solution to both problems was presented by Fischer and Almlöf [12], who suggested the use of a quasi-Newton approach using an approximate hessian to avoid its accurate computation, and also devised a very efficient recursive algorithm which allows the updating of the inverse hessian multiplied by gradient vectors without actual storage of the hessian matrix. Another solution to both problems has been given by Shepard [13], but the accurate computation of the hessian requires additional builds of Fock-like matrices, and thus a considerable increase in computational cost [16].

The purpose of this paper is to demonstrate how an approximate second order method based on the hessian update procedure suggested by Fischer and Almlöf can be used to perform efficient orbital optimizations, not only for closed shell restricted Hartree-Fock (RHF) [12], but also for other types of wavefunctions, namely restricted open shell Hartree-Fock (ROHF), GVB, and MCSCF. These procedures have been implemented in the electronic structure code GAMESS [17], and are the present default convergence methods.

## 2 Closed shell RHF wavefunctions

Second order SCF (SOSCF) methods for RHF wavefunctions have been described in previous papers [5–16]. Here, we outline this method and present some details of our implementation of it.

The closed shell RHF wavefunction can be represented as an anti-symmetrized product of doubly occupied orbitals  $\psi_i$ :

Correspondence to: M. S. Gordon

$$\Psi_{RHF} = \hat{A} \left[ \psi_1(1) \bar{\psi}_1(2) \dots \psi_{N/2}(N-1) \bar{\psi}_{N/2}(N) \right] \quad (1.1)$$

where  $N$  is the number of electrons. The electronic energy is given by

$$E = \sum_i^{occ} 2h_{ii} + \sum_{i,j}^{occ} (2J_{ij} - K_{ij}) \quad (1.2)$$

$$\text{where } h_{ii} = \int \psi_i(1) \mathbf{h}(1) \psi_i(1) dV_1 \quad (1.3)$$

$$J_{ij} = (ii|jj) = \iint \psi_i(1) \psi_j(2) \frac{1}{r_{12}} \psi_i(1) \psi_j(2) dV_1 dV_2 \quad (1.4)$$

$$K_{ij} = (ij|ij) = \iint \psi_i(1) \psi_j(2) \frac{1}{r_{12}} \psi_j(1) \psi_i(2) dV_1 dV_2. \quad (1.5)$$

The variational condition leads to the familiar HF equations

$$\mathbf{F}\psi_i = \varepsilon_i \psi_i \quad (1.6)$$

where  $\mathbf{F}$  is the Fock operator:

$$\mathbf{F} = \mathbf{h} + \sum_j^{occ} (2\mathbf{J}_j - \mathbf{K}_j). \quad (1.7)$$

The standard method of solution of the HF equations includes constructing the Fock matrix in the atomic orbital (AO) basis, transforming it to the current molecular orbital (MO) basis, and diagonalizing. Diagonalization gives new MOs which are used for the next iteration, and this is repeated until convergence.

In the second order method based on exponential parametrization [8, 9], the new set of orbitals is obtained from the old by an orthogonal transformation represented as an exponential of an antisymmetric matrix:

$$\mathbf{C}_{new} = \mathbf{C}_{old} \mathbf{U} = \mathbf{C}_{old} \exp(\mathbf{A}), \quad (1.8)$$

$$\text{where } \exp(\mathbf{A}) = \mathbf{I} + \mathbf{A} + 1/2 \mathbf{A}^2 + \dots \quad (1.9)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{x} \\ -\mathbf{x} & \mathbf{0} \end{bmatrix}. \quad (1.10)$$

The  $x_{ia}$  elements are the  $L = n_{occ} \times n_{virt}$  independent rotational parameters, each corresponding to a rotation between occupied orbital  $i$  and virtual orbital  $a$ . The variational condition is:

$$\frac{\partial E}{\partial x_{ia}} = g_{ia} = 0 \quad (1.11)$$

where  $g_{ia}$  are the elements of the orbital gradient of the energy. This equation can be solved using the Newton-Raphson method:

$$\delta_n = \mathbf{x}_{n+1} - \mathbf{x}_n = -\mathbf{B}_n^{-1} \cdot \mathbf{g}_n \quad (1.12)$$

where  $\delta_n$  is the displacement vector,  $\mathbf{g}_n$  is the gradient vector, and  $\mathbf{B}_n$  is the hessian matrix on iteration  $n$ .

Using the Newton-Raphson method in a fixed coordinate system requires very time-consuming computations of the first and second derivatives of the energy with respect to the parameters  $x_{ia}$ . The first derivatives  $\frac{\partial E}{\partial x_{ia}}$  can be calculated exactly (see Ref. [14] for exact formulae), but the calculations involve many matrix operations. Our implementation, like most others, does

not use fixed coordinates. Instead, we set the parameters  $x_{ia}$  to zero at the beginning of each iteration and use the new displacement  $\delta_n = \mathbf{x}_{n+1}$  to update the orbital coefficient matrix obtained on the previous iteration  $\mathbf{C}_{n+1} = \mathbf{C}_n \exp(\mathbf{A}_{n+1})$ . This gives much simpler expressions for the derivatives at  $\mathbf{x} = \mathbf{0}$ . For example, the first derivative [8] is simply:

$$g_{ia} = \left. \frac{\partial E}{\partial x_{ia}} \right|_{x_{ia}=0} = 4 \langle \psi_i | \mathbf{F} | \psi_a \rangle = 4F_{ia} \quad (1.13)$$

where  $F_{ia}$  is an element of the Fock matrix transformed into the MO basis. The orthogonal transformation matrix  $\mathbf{U} = \exp(\mathbf{A})$  is obtained by truncating the expansion to first order:

$$\exp(\mathbf{A}) = \mathbf{I} + \mathbf{A} \quad (1.14)$$

followed by a Schmidt orthogonalization. We have found that addition of the second order term  $1/2\mathbf{A}^2$  does not lead to any decrease in the total number of iterations. Other more exact expressions for the rotation matrix, such as the Klein-Cayley formula

$$\mathbf{U} = (\mathbf{I} + \mathbf{A}/2)/(\mathbf{I} - \mathbf{A}/2) \quad (1.15)$$

have not been considered due to our desire to minimize the number of floating point operations to obtain  $\mathbf{U}$ .

The second derivative expressions at  $\mathbf{x} = \mathbf{0}$  [8] include not only Fock matrix elements, but also two-electron integrals in the MO basis. To avoid the expensive integral transformation required to compute the exact hessian matrix, we follow Fischer and Almlöf [12] and start with an approximate diagonal hessian, which thus can be inverted trivially. The inverse orbital hessian is then updated using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) updating formula [18] via the recursive update recipe given in Ref. [12]. The greatest advantage of this procedure is memory savings, since the hessian matrix itself does not need to be kept in memory. The recursive algorithm requires storage of three vectors of size  $L$  (where  $L$  is the number of rotational parameters) from each previous iteration. We write these vectors to disk and read them on each iteration to update hessian related information.

Our experience shows that the choice of the approximate initial diagonal hessian is very important to provide convergence for this method. The use of a unit matrix for this purpose is certainly not satisfactory. Consideration of the general formula for the RHF orbital hessian [8] ( $i,j$  = occupied orbitals;  $a,b$  = virtual orbitals)

$$B_{i_a, j_b} = 4F_{ab} \delta_{ij} - 4F_{ij} \delta_{ab} + 4[4(ia|bj) - (ja|ib) - (ab|ij)] \quad (1.16)$$

suggests that the use of orbitals which are approximately canonical will increase the diagonal dominance of the hessian. Since our Huckel guess orbitals [17] are not nearly canonical, we diagonalize the Fock matrix on the first iteration, and on the next cycle take as our approximate initial diagonal hessian

$$B_{i_a, i_a} = 4F_{aa} - 4F_{ii} = 4(\varepsilon_a - \varepsilon_i) \quad (1.17)$$

**Table 1.** Comparison of second order SCF and DIIS methods<sup>a</sup>

Molecule	No. of basis functions	No. of iterations		SCF time (s)		Fock time (s)	
		DIIS	SOSCF	DIIS	SOSCF	DIIS	SOSCF
Glycine	85	12	12	3.1	0.8	355	Same as DIIS
Thymine	149	13	13	17.4	5.7	2,243	Same as DIIS
Nicotine	208	13	12	48.4	14.1	6,361	6,123
Luciferin	294	14	15	153.6	51.1	13,733	14,574
Cyclic AMP	356	14	13	275.4	77.4	24,262	23,647

<sup>a</sup> The calculations were performed on a 67 MHz IBM SP2 thin node running sequentially. The basis set is 6-31 G(d), and none of the molecules possesses any symmetry. All runs used direct SCF techniques

**Table 2.** Number of SCF iterations during geometry optimization of glycine molecule

Geometry step	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
DIIS	13	11	11	10	10	8	8	8	8	8	9	9	7	7	7	6
SOSCF	13	10	10	9	8	7	7	7	6	7	8	8	6	6	5	5

**Fig. 1.** Summary of approximate second order SCF algorithm, with FLOP count<sup>a</sup>

STEP	COST
0. Estimate $\mathbf{B}^{-1}$ , on first iteration <sup>b</sup>	2L
1. Obtain gradient in MO basis <sup>b</sup>	$2n_{occ} N^2 + 2LN$
2. Apply BFGS updates to obtain the orbital rotation parameters $\mathbf{x} = \mathbf{B}^{-1} \mathbf{g}$	$16L + 20(n_{iter-2})L$
3. Generate transformation $\mathbf{U} = \mathbf{I} + \mathbf{A}$ from $\mathbf{x}$	negligible
4. Schmidt orthogonalize $\mathbf{U}$	$2N^3 + N^2$
5. Rotate orbitals, $\mathbf{C}_{n+1} = \mathbf{C}_n \mathbf{U}$	$2N^3$

<sup>a</sup> N is the size of the AO basis, L is the number of independent orbital rotation parameters, and  $n_{iter}$  is the number of iterations since SOSCF was initiated. <sup>b</sup> The costs of these steps are wavefunction dependent, and values shown are for RHF computations

since the two-electron integral terms from (1.16) are much smaller than the Fock contributions. At subsequent geometries, it suffices to use orbital energies from the previous geometry to form the initial diagonal orbital hessian.

The entire orbital update algorithm is shown in Fig. 1. We begin the algorithm when the maximum orbital gradient element falls below a threshold, normally 0.25. Our Huckel guess orbitals [17] normally meet this criterion after the conventional diagonalization on the first iteration, but if the starting orbitals are very poor we rely on conventional extrapolation to bring us within the radius of convergence of our SOSCF method. The SCF process is considered converged when the maximum orbital gradient component is one microhartree. This is roughly equivalent to  $10^{-5}$  convergence of the density matrix. Since applications such as nuclear gradient and hessian computation, perturbation theory, and Koopmans' theorem require canonical orbitals, we always perform one Fock matrix diagonalization after final convergence.

The convergence of this approximate second order method is not, of course, quadratic because of the various approximations used, but it is superlinear and is comparable with that of the DIIS method [2]. However, the number of computations is less for the approximate

second order method than for DIIS. The floating point operation (FLOP) count of our RHF DIIS implementation [17] is roughly  $17 N^3$ , from seven matrix multiplies ( $14 N^3$ ) and one diagonalization ( $10/3 N^3$ ). Taking  $n_{occ} = 1/3 N$  as typical of RHF calculations, Fig. 1 shows that the approximate SOSCF orbital improvement requires roughly  $5N^3$  FLOPs.

Table 1 shows several examples of organic molecules converged with the conventional Fock diagonalization method + DIIS convergence accelerator, compared to the SOSCF method described above. The two final columns show the times needed for computation of integrals and construction of the Fock matrix in direct mode. As can be seen from the table, the number of iterations is practically the same for both methods ( $\pm 1$  iteration), but the time required to solve the HF equations is about 3 times less for SOSCF than for DIIS. This is not very important for a serial run, since much more time is required to calculate integrals and form Fock matrices than to solve the HF equations. However, if integrals are calculated in parallel (assuming perfect parallelization of the integrals and Fock build), then the smaller time of the sequential solve step gives better results for speedups. For the last example of cyclic AMP, the theoretical speedup on 32 nodes,

$$\text{DIIS: } (24,262 + 275)/(24,262/32 + 275) = 23.7$$

$$\text{SOSCF: } (23,647 + 77)/(23,647/32 + 77) = 29.0$$

is much better in the case of SOSCF.

Another advantage of using this SOSCF method is a decreased number of iterations during geometry optimizations. Table 2 presents the number of SCF iterations on each geometry optimization step for glycine. The number of iterations is consistently 1-2 iterations less for SOSCF on all subsequent geometries. This saves a total of 18 iterations during this optimization run.

### 3 High spin open shell ROHF wavefunctions

The high spin coupled restricted open shell HF wavefunction can be written as:

**Table 3.** Comparison of different GVB orbital optimization methods (number of iterations)

Molecule	OCBSE	GVB-DIIS <sup>a</sup>	SOSCF
Methylene, CH <sub>2</sub> , 1P <sup>b</sup>	10	11	10
Bicyclobutane, C <sub>4</sub> H <sub>6</sub> , 1P	Does not converge	23	18
Disilane, Si <sub>2</sub> H <sub>6</sub> , 1P	Does not converge	Does not converge	10
Ethylene, C <sub>2</sub> H <sub>4</sub> , 2P	12	11	11
Stannobenzene, 3P	46	Does not converge	31
Cyclic N <sub>2</sub> O <sub>2</sub> , 5P	23	Does not converge	14

<sup>a</sup>These results are given for Single Fock Operator + DIIS methods implemented in GAMESS (single-pair GVB-DIIS<sup>2c</sup> and multi-pair GVB-DIIS<sup>3</sup>)

<sup>b</sup>The notation nP indicates the number of GVB pairs used

$$\Psi_{ROHF} = \hat{A}[\psi_1\bar{\psi}_1\psi_2\bar{\psi}_2\dots\dots\psi_{N-M}\bar{\psi}_{N-M}\psi_{N-M+1}\dots\dots\psi_N] = \hat{A}[\Psi_{core}\Psi_{open}] \quad (2.1)$$

where  $N - M$  is the number of doubly occupied (closed shell) orbitals, and  $M$  is the number of singly occupied (open shell) orbitals.

Previously, high spin ROHF wavefunctions were optimized in GAMESS by diagonalizing a single Fock matrix in the MO basis, constructed as follows [17]:

$$\mathbf{F}_{ROHF} = \begin{bmatrix} F_2 & F_\beta & (F_\alpha + F_\beta)/2 \\ F_\beta & F_1 & F_\alpha \\ (F_\alpha + F_\beta)/2 & F_\alpha & F_0 \end{bmatrix} \quad (2.2)$$

where  $F_\alpha$  and  $F_\beta$  are alpha and beta Fock matrices transformed to the MO basis.  $F_2$ ,  $F_1$ , and  $F_0$  are doubly, singly, and zero occupied diagonal blocks:

$$F_2 = A_{cc}F_\alpha + B_{cc}F_\beta \quad (2.3a)$$

$$F_1 = A_{oo}F_\alpha + B_{oo}F_\beta \quad (2.3b)$$

$$F_0 = A_{vv}F_\alpha + B_{vv}F_\beta. \quad (2.3c)$$

The best convergence rate is found with Roothaan's A and B canonicalization coefficients [19].

To use our quasi-Newton approach for ROHF wavefunctions, we construct the orbital gradient vector from the off-diagonal blocks of the above Fock matrix:

$$g_{ik} = F_{ik}^\beta \quad \text{if } i = \text{doubly occupied,} \\ \quad \quad \quad \quad \quad \quad \quad \quad k = \text{singly occupied orbital} \quad (2.4a)$$

$$g_{ia} = \frac{1}{2}(F_{ia}^\alpha + F_{ia}^\beta) \quad \text{if } i = \text{doubly occupied,} \\ \quad \quad \quad \quad \quad \quad \quad \quad a = \text{virtual orbital} \quad (2.4b)$$

$$g_{ka} = F_{ka}^\alpha \quad \text{if } k = \text{singly occupied,} \\ \quad \quad \quad \quad \quad \quad \quad \quad a = \text{virtual orbital.} \quad (2.4c)$$

As in the case of RHF, we diagonalize the above Fock matrix on the very first iteration and use its eigenvalues  $\epsilon$  to approximate the initial diagonal hessian:

$$B_{ik,ik} = \epsilon_k - \epsilon_i \quad (2.5a)$$

$$B_{ia,ia} = \epsilon_a - \epsilon_i \quad (2.5b)$$

$$B_{ka,ka} = \epsilon_a - \epsilon_k \quad (2.5c)$$

where the index  $i$  denotes doubly occupied,  $k$  singly occupied, and  $a$  virtual orbitals. The rest of our implementation is the same as for closed shell RHF wavefunctions.

## 4 GVB and low spin ROHF wavefunctions

A GVB wavefunction can be represented as [3, 20]:

$$\Psi_{GVB} = \hat{A}[\Psi_{core}\Psi_{open}\Psi_{pair}] \quad (3.1)$$

where  $\Psi_{core}$  is a product of doubly occupied orbitals,  $\Psi_{open}$  is a product of singly occupied orbitals, and

$$\Psi_{pair} = \prod_{i=1}^{n_{pair}} (C_{gi}\phi_{gi}(1)\phi_{gi}(2) + C_{ui}\phi_{ui}(1)\phi_{ui}(2))(\alpha(1)\beta(2) - \beta(1)\alpha(2)). \quad (3.2)$$

Here,  $\phi_{gi}$  and  $\phi_{ui}$  are orthogonal GVB natural orbitals, and  $C_{gi}$  and  $C_{ui}$  are GVB configuration interaction (CI) coefficients. This wavefunction leads to the energy expression [20]:

$$E = \sum_i^{occ} 2f_i h_{ii} + \sum_{i,j}^{occ} (a_{ij}J_{ij} + b_{ij}K_{ij}) \quad (3.3)$$

$$\text{where } h_{ii} = \langle \phi_i | \mathbf{h} | \phi_i \rangle \quad (3.4)$$

$$J_{ij} = (ii|jj) = \int \phi_i(1)\mathbf{J}^j(1)\phi_i(1)d\nu_1 \quad (3.5)$$

$$K_{ij} = (ij|ij) = \int \phi_i(1)\mathbf{K}^j(1)\phi_i(1)d\nu_1 \quad (3.6)$$

$\mathbf{h}$ ,  $\mathbf{J}^j(1)$ , and  $\mathbf{K}^j(1)$  are one- and two-electron (Coulomb and exchange) operators, and  $f_i$ ,  $a_{ij}$ ,  $b_{ij}$  are orbital occupation coefficients (see Refs. [3, 20] for their definitions). The variational condition using this energy expression is:

$$\delta E = \sum_i \langle \delta \phi_i | \mathbf{F}^i | \phi_i \rangle = 0 \quad (3.7)$$

where  $\mathbf{F}^i$  is the Fock operator for the orbital  $\phi_i$ :

$$\mathbf{F}^i = f_i \mathbf{h} + \sum_j^{occ} (a_{ij}\mathbf{J}^j + b_{ij}\mathbf{K}^j). \quad (3.8)$$

Orbitals with the same occupation coefficients, and therefore the same Fock operator, form a shell.

Several methods have been used to optimize orbitals of GVB wavefunctions. The standard OCBSE (orthogonality constrained basis set expansion) procedure is based on diagonalizing the Fock matrices corresponding to each shell [20]. An orbital DIIS method has been suggested to improve convergence of this method [4]. Another approach is to diagonalize a single combined Fock matrix. Several possible generalized Fock opera-

tors are discussed in Ref. [3] together with the use of DIIS with this approach. Although their suggested GVB-DIIS Fock operator [3] introduces second-order mixing, we find that our quasi-Newton optimization based on the hessian update method of Fisher and Almlöf [12] works more reliably and in fewer iterations.

The general formula for the GVB orbital gradient (that has to be brought to zero at convergence) is:

$$g_{ij} = \frac{\partial E}{\partial x_{ij}} = 4F_{ij}^J - 4F_{ij}^I \quad (3.9)$$

which reduces to:

$$g_{ij} = 4F_{ij}^J \quad \text{if } j = \text{occupied orbital from shell } J, \\ i = \text{virtual orbital} \quad (3.10)$$

$$g_{ij} = 0 \quad \text{if orbitals } i \text{ and } j \text{ belong to} \\ \text{the same shell} \quad (3.11)$$

$$g_{ij} = 4F_{ij}^J - 4F_{ij}^I \quad \text{if } i \text{ and } j \text{ are occupied orbitals} \\ \text{from different shells } I \text{ and } J. \quad (3.12)$$

The general formula [3, 8, 21] for the diagonal elements of the GVB orbital hessian in the MO basis is:

$$B_{ij,ij} = \frac{\partial^2 E}{\partial x_{ij}^2} = 4(F_{ii}^J - F_{ii}^I) - 4(F_{jj}^J - F_{jj}^I) \\ + 8(a_{ii} + a_{jj} - 2a_{ij})(ij|ij) + 4(b_{ii} + b_{jj} - 2b_{ij}) \\ [(ii|jj) + (ij|ij)]. \quad (3.13)$$

In the case for which  $i$  is a virtual orbital,  $F^I = 0$ ,  $a_{ii} = b_{ii} = 0$ ,  $a_{ij} = b_{ij} = 0$ , and  $B_{ij,ij}$  becomes:

$$B_{ij,ij} = 4F_{ii}^J - 4F_{jj}^J + 4(2a_{jj} + b_{jj})(ij|ij) + 4b_{jj}(ii|jj) \quad (3.14)$$

Since the  $(ii|jj) = J_{ij}$  and  $(ij|ij) = K_{ij}$  matrices corresponding to open shell  $\mathbf{J}$  and  $\mathbf{K}$  operators are already

available in the MO basis as a by-product of the GVB Fock matrix formation, we can calculate all diagonal hessian elements exactly, except those that correspond to core-virtual rotations. In the case for which  $j$  is doubly occupied and  $i$  is a virtual orbital ( $a_{jj} = 2, b_{jj} = -1$ ):

$$B_{ij,ij} = 4F_{ii}^J - 4F_{jj}^J + 4(3(ij|ij) - (ii|jj)). \quad (3.15)$$

It would require additional integral transformation work to get the two-electron integrals in (3.15), and therefore we neglect these terms. Fortunately, the neglected terms are much smaller than the remaining part

$$B_{ij,ij} = 4F_{ii}^J - 4F_{jj}^J \quad (3.16)$$

which is the same expression we use in the case of closed shell RHF. Note that all terms are, however, important for hessian elements corresponding to core-pair, pair-pair, and pair-virtual rotations, and these are calculated exactly. As in the case of RHF, we perform one single combined Fock matrix diagonalization on the first iteration to provide a better initial orbital hessian for our approximate second order (SOSCF) method.

Several examples of GVB calculations are given in Table 3. The middle column of the table represents the original one-pair GVB-DIIS<sup>2c</sup> implemented in GAMESS several years ago, while the multi-pair results are obtained from our implementation of the GVB-DIIS multishell Fock operator method given in Ref. [3]. GVB wavefunctions can be difficult to converge, and each of the three methods fails from time to time. We find, however, that the approximate second order method is the most successful in converging the majority of GVB cases we have tried.

The energy formula (3.3), as well as the gradient (3.9) and hessian (3.13) formulae can be also used for different kinds of low spin ROHF calculations by using the appropriate coefficients  $f_i, a_{ij}$ , and  $b_{ij}$ . GAMESS contains values of these coefficients for some cases of singly occupied orbitals, and they can be input for other cases of partial orbital occupancies. Both the DIIS method based on a single combined Fock matrix and the approximate second order method work very successfully for all kinds of low spin ROHF calculations. Two examples are shown in Table 4.

**Table 4.** Comparison of different methods for low-spin ROHF calculations

Molecule	OCBSE	DIIS	SOSCF
H <sub>2</sub> CO n → π* <sup>1</sup> A''	Does not converge	10	11
Y atom s <sup>2</sup> d <sup>1</sup> <sup>2</sup> D	Does not converge	8	9

**Table 5.** Comparison of three MCSCF orbital improvement methods<sup>a</sup>

Molecule <sup>b</sup>	2nd order method		1st order method		Approx. 2nd order	
	Iterations	Time (s)	Iterations	Time (s)	Iterations	Time (s)
SiH <sub>2</sub>	7	18	23	25	12	13
N <sub>3</sub> F	8	333	31	270	16	145
C <sub>4</sub> H <sub>4</sub>	7	1398	12	496	12	441
CH <sub>3</sub> AsNCH <sub>3</sub>	9	5608	34	2432	14	1194
7-azaindole	–	–	14	15,018	9	10,098

<sup>a</sup> CPU timings for these runs are taken on 42 MHz model 350 IBM RS/6000 workstation

<sup>b</sup> The MCSCF (n, m) examples are complete active spaces with n electrons in m orbitals: (1) MCSCF(6,6)/6-31G (d, p) calculation of triplet SiH<sub>2</sub> (29 AOs, 51 CSFs); (2) MCSCF(6,6)/6-31G(d) calculation of cyclic isomer of N<sub>3</sub>F (60 AOs, 92 CSFs); (3) MCSCF (4,4)/6-31 ++ G (d, p) ground state of C<sub>4</sub>H<sub>4</sub> (100 AOs, 12 CSFs); (4) MCSCF(4,4)/DZP for CH<sub>3</sub>-As=N-CH<sub>3</sub> compound (103 AOs, 12 CSFs); (5) MCSCF(10,9)/DZP for 7-azaindole (165 AOs, 5292 CSFs). The job is impossible to run with the exact second order program because of memory requirements

## 5 MCSCF wavefunctions

The approximate second order update method is also implemented for the orbital improvement step for MCSCF wavefunction optimization. Space precludes a full review of all previously proposed MCSCF optimization schemes here. Early first order methods based on the Generalized Brillouin Theorem [22–24] were abandoned by most workers with the advent of second-order Newton-Raphson methods [25–27]. Recently a Renormalized Fock Operator approach based on diagonalization of a matrix of single excitation matrix elements has been proposed by Meier and Staemmler [28]. This method exhibits first order convergence, and has recently been implemented in HONDO8 by Dupuis et al. [29]. This code served as the basis of our approximate second order MCSCF scheme.

The MCSCF wavefunction  $\Psi$  is a superposition of electronic configurations  $\Phi_K$ :

$$\Psi_{MCSCF} = \sum_K A_K \Phi_K, \quad (4.1)$$

where  $A_K$  are the CI coefficients. Typically, many of the orbitals can be constrained to be doubly occupied in all configurations – these are *core* (or inactive) orbitals and will be denoted  $i, j, k$ . *Active* orbitals are allowed to have variable occupation numbers and are denoted  $t, u, v, w$ . *Virtual* orbitals are those that are empty in all configurations, and are denoted  $a, b, c$ . We use the notation  $p, q, r, s$  for general orbitals.

The energy expression for an MCSCF wavefunction is:

$$E = \sum_{pq} \gamma_{pq} h_{pq} + \frac{1}{2} \sum_{pqrs} \Gamma_{pqrs} (pq|rs) \quad (4.2)$$

where  $\gamma_{pq}$  and  $\Gamma_{pqrs}$  are the elements of one- and two-body density matrices, depending on  $A_K$ . The one- and two- electron integrals  $h_{pq}$  and  $(pq|rs)$  depend on the orbitals. The optimization of CI coefficients and molecular orbitals may be separated, or unfolded, into two steps: solving the secular equation for  $A_K$ , followed by some sort of orbital improvement scheme.

The first order method suggested for MCSCF orbital improvement by Meier and Staemmler [28] and implemented in HONDO by Dupuis et al. [29] has many advantages. Although it requires more iterations to converge than second order (Newton-Raphson) methods, the time needed for each iteration is much less. Each macroiteration in the first order algorithm includes several microiterations which skip the CI step and part of the integral transformation. Each microiteration [29] consists of diagonalization of an approximate Fock matrix which is formed using old density matrices and partially updated two-electron integrals. The fairly numerous microiterations in the first order scheme add a modest, but non-negligible, amount of time to each

macroiteration. In contrast, the high cost of the full second order method comes from the construction of the orbital hessian matrix which requires  $(pt|qu)$  and  $(pq|tu)$  types of molecular integrals with two indices  $p, q$  running over the entire orbital space. Only a single general orbital index needs to be transformed in order to construct the Lagrangian matrix for the first order MCSCF. Of course, another problem associated with using second order methods is storage of the orbital hessian matrix.

The method we have implemented in GAMESS is an attempt to use all the advantages of the first order MCSCF method, while also introducing some features of the second order approach. Again, we construct an approximate initial diagonal hessian and use the update procedure suggested by Fisher and Almlöf [12]. The resulting convergence behavior is not, of course, as good as that of an exact Newton-Raphson method, but is better than that of a first order method.

Our implementation of the approximate second order method in GAMESS uses part of the first order MCSCF program written by Dupuis for HONDO [29]. The Lagrangian matrix  $G$  calculated by the first order code is used to construct the orbital gradient vector. The components of the orbital gradient corresponding to core-virtual (a), active-virtual (b), and core-active (c) rotations are:

$$g_{ia} = \frac{\partial E}{\partial x_{ia}} = 4(F_{ai}^{core} + F_{ai}^{act}) = G_{ai} \quad (4.3a)$$

$$g_{ta} = \frac{\partial E}{\partial x_{ta}} = 2 \sum_u \gamma_{tu} F_{au}^{core} + 2 \sum_{u,v,w} \Gamma_{tuvw} (au|vw) = G_{at} \quad (4.3b)$$

$$g_{it} = \frac{\partial E}{\partial x_{it}} = 4(F_{ii}^{core} + F_{ii}^{act}) - 2 \left[ \sum_u \gamma_{tu} F_{iu}^{core} + \sum_{u,v,w} \Gamma_{tuvw} (iu|vw) \right] = G_{ii} - G_{it} \quad (4.3c)$$

where  $\gamma$  and  $\Gamma$  are the one- and two-body density matrices, and  $F^{core}$  and  $F^{act}$  are two Fock type matrices:

$$F_{pq}^{core} = h_{pq} + \sum_k^{core} [2(pq|kk) - (pk|qk)] \quad (4.4)$$

$$F_{pq}^{act} = \sum_{u,v} \gamma_{uv} \left[ (pq|uv) - \frac{1}{2}(pu|qv) \right]. \quad (4.5)$$

Explicit expressions for the MCSCF orbital hessian in terms of two-electron integrals and density matrix elements are given in Ref. [26]. The exact formulae for diagonal orbital hessian elements are:

$$B_{ia,ia} = \frac{\partial^2 E}{\partial x_{ia}^2} = 4(F_{aa}^{core} + F_{aa}^{act}) - 4(F_{ii}^{core} + F_{ii}^{act}) + 4[3(ai|ai) - (aa|ii)] \quad (4.6a)$$

$$\begin{aligned}
B_{ta,ta} &= \frac{\partial^2 E}{\partial x_{ta}^2} = 2\gamma_{tt} F_{aa}^{core} - 2 \sum_u \gamma_{tu} F_{tu}^{core} \\
&\quad - 2 \sum_{u,v,w} \Gamma_{tuvw}(tu|vw) \\
&\quad + 2 \sum_{u,v} \{\Gamma_{ttw}(aa|uv) + 2\Gamma_{vtu}(av|au)\}
\end{aligned} \tag{4.6b}$$

$$\begin{aligned}
B_{it,it} &= \frac{\partial^2 E}{\partial x_{it}^2} = 4(F_{tt}^{core} + F_{tt}^{act}) - 4(F_{ii}^{core} + F_{ii}^{act}) \\
&\quad + 2\gamma_{tt} F_{ii}^{core} - 2 \sum_u \gamma_{tu} F_{tu}^{core} - 2 \sum_{u,v,w} \Gamma_{tuvw}(tu|vw) \\
&\quad + 2 \sum_{u,v} \{\Gamma_{ttv}(uv|ii) + 2\Gamma_{vtu}(ui|vi)\} \\
&\quad + 4 \sum_u \{(\delta_{tu} - \gamma_{tu})[3(ui|ti) - (tu|ii)]\}.
\end{aligned} \tag{4.6c}$$

No active-active hessian elements are shown, as we restrict our method to complete active spaces. In the framework of the first order MCSCF approach [29], an integral transformation is performed only for  $F^{core}$ ,  $F^{act}$ , and  $(pu|vw)$  types of integrals with only one general index  $p$  running through all orbital space, and three other indices over active orbitals only. Accordingly, we ignore those terms of the above expressions that contain two-electron integrals with two non-active indices. The truncated diagonal hessian contains only  $F^{core}$  and  $F^{act}$  matrix elements and  $(tu|vw)$  integrals with all indices in the active space. This approximation is satisfactory for  $B_{ia,ia}$  hessian elements, since the last term is small compared with the first two. In the case of  $B_{ta,ta}$  and  $B_{it,it}$  elements, the last terms may be quite substantial. Ignoring these terms may result in negative signs for some of the hessian elements. This causes severe convergence problems when starting a Newton-Raphson optimization from such an initial hessian.

To improve this situation, we add  $2\gamma_{tt} F_{aa}^{act}$  to  $B_{ta,ta}$  and  $2\gamma_{tt} F_{ii}^{act}$  to  $B_{it,it}$  hessian elements. In the limit of active orbital  $t$  becoming weakly occupied ( $\gamma_{tt} \rightarrow 0$ ), the final term of (4.6b) and  $2\gamma_{tt} F_{aa}^{act}$  both vanish. In the limit of active orbital  $t$  becoming nearly filled ( $\gamma_{tt} \rightarrow 2$ ,  $\Gamma_{ttw} \rightarrow 2\gamma_{tw}$ ,  $\Gamma_{ttv} \rightarrow -\frac{1}{2}\gamma_{tv}$ ),  $2\gamma_{tt} F_{aa}^{act}$  becomes identical to the final term of (4.6b). Similarly, in the limit  $t \rightarrow$  doubly occupied, the two final terms in (4.6c) reduce to the same expression as  $2\gamma_{tt} F_{ii}^{act}$ . The  $2\gamma_{tt} F_{ii}^{act}$  terms prove to be a satisfactory compensation for the exact terms and provide a positive definite starting hessian for use with the quasi-Newton update procedure. The resulting diagonal hessian used to start the approximate second order MCSCF orbital improvement is therefore calculated as follows:

$$B_{ia,ia} = 4(F_{aa}^{core} + F_{aa}^{act}) - 4(F_{ii}^{core} + F_{ii}^{act}) \tag{4.7a}$$

$$\begin{aligned}
B_{ta,ta} &= 2\gamma_{tt} F_{aa}^{core} - 2 \sum_u \gamma_{tu} F_{tu}^{core} - 2 \sum_{u,v,w} \Gamma_{tuvw}(tu|vw) \\
&\quad + 2\gamma_{tt} F_{aa}^{act}
\end{aligned} \tag{4.7b}$$

$$\begin{aligned}
B_{it,it} &= 4(F_{tt}^{core} + F_{tt}^{act}) - 4(F_{ii}^{core} + F_{ii}^{act}) + 2\gamma_{tt} F_{ii}^{core} \\
&\quad - 2 \sum_u \gamma_{tu} F_{tu}^{core} - 2 \sum_{u,v,w} \Gamma_{tuvw}(tu|vw) + 2\gamma_{tt} F_{ii}^{act}.
\end{aligned} \tag{4.7c}$$

A previous attempt to reduce the amount of work in the integral transformation by approximating the orbital hessian has been made by Camp et al. [30]. These workers made the same approximation of dropping the final term of (4.6a), but calculated sufficient integrals with two virtual indices to evaluate the final terms of (4.6b) and (4.6c). Our approach is to transform only one virtual index, requiring the additional approximations just described.

Since very little additional work is required to construct the approximate diagonal hessian and it is done only on the initial iteration, the cost of each quasi-Newton iteration is less than that of one first order macroiteration, which consists of several Fock matrix diagonalizations (microiterations) [28, 29]. The convergence rate is also better than that of the first order method. As a result, the approximate second order method is very fast and useful for MCSCF calculations of very large molecular systems.

Table 5 shows several examples [31] of MCSCF calculations using three methods for orbital improvement: (1) an exact second order method with the full orbital hessian matrix calculated on each iteration; (2) a first order MCSCF method based on effective Fock matrix diagonalization, supplemented by DIIS convergence acceleration [29]; (3) the approximate second order method described above. As can be seen from Table 5, the approximate second order approach provides substantial savings of computer time. Note that the final example is a MCSCF(10,9) calculation of 7-azaindole ( $C_7N_2H_6$ , 165 AOs) which proved impossible to do with the exact second order program because insufficient memory was available. About 22,000,000 words of memory are required for the orbital update procedure with the exact Newton-Raphson scheme, but fewer than 1,000,000 words are needed in the case of the approximate quasi-Newton approach. The exact second order method is, however, preferable for MCSCF calculations with large numbers of configurations, but small basis sets, since in this case the CI part (solution for  $A_K$ 's) is more time consuming than the orbital improvement part, and the least number of iterations is desirable.

## 6 Conclusion

We have demonstrated an efficient way to optimize molecular orbitals for different types of wavefunctions (RHF, ROHF, GVB, and MCSCF) by extending the approximate second order method with a diagonal hessian update procedure due to Fisher and Almlöf. The algorithm is quite simple, and it is remarkable that essentially the same process can be used to optimize all these classes of wavefunctions. Details relating to our implementation, and explicit formulae for the exact

orbital gradient and the approximate diagonal orbital hessian have been presented for each kind of wavefunction. The method is faster than the standard diagonalization techniques used for RHF and GVB. Its convergence properties (number of iterations) are comparable to that of DIIS accelerated SCF methods, but the time needed for solving the HF equations is about three times less. In addition, the method eliminates the traditional diagonalization step (except for the final orbital canonicalization) which is not easily parallelizable. For MCSCF, the approximate second order orbital update is trivial compared to the cost of the orbital gradient, and convergence rates are intermediate between the Renormalized Fock Operator first order method and the exact second order approach.

*Acknowledgements.* We thank Michel Dupuis, who provided us with his first order MCSCF code from HONDO8, and both Michel Dupuis and Klaus Ruedenberg for helpful comments upon the manuscript. Funding for this work was provided by the Air Force Office of Scientific Research (grant F49620-95-1-0077).

## References

1. Roothaan CCJ (1951) *Rev Mod Phys* 23: 69
2. a) Pulay P (1980) *Chem Phys Lett* 73: 393; b) Pulay P (1982) *J Comp Chem* 3: 556; c) Hamilton TP, Pulay P (1986) *J Chem Phys* 84: 5728; d) Sellers H (1991) *Chem Phys Lett* 180: 461; e) Sellers H (1993) *Int J Quantum Chem* 45: 31
3. Muller RP, Langlois J-M, Ringnalda M, Friesner RA, Goddard III, WA (1994) *J Chem Phys* 100: 1226
4. a) Ionova, IV, Carter, EA (1995) *J Chem Phys* 102: 1251; b) Ionova IV, Carter EA (1996) *J Comput Chem* 17: 1836
5. Thouless DJ (1960) *Nucl Phys* 21: 225
6. a) Brown TH, Taylor RL (1965) *J Phys Chem* 69: 2316; b) Brown TH (1968) *J Chem Phys* 49: 2291
7. Wessel WR (1967) *J Chem Phys* 47: 3253
8. Douady J, Ellinger Y, Subra R, Levy B (1980) *J Chem Phys* 72: 1452
9. a) Bacskay GB (1981) *Chem Phys* 61: 385; b) Bacskay GB (1982) *Chem Phys* 65: 383
10. Head-Gordon M, Pople JA (1988) *J Phys Chem* 92: 3063
11. Sano T, I'Haya YJ (1991) *J Chem Phys* 95: 6607
12. Fischer TH, Almlöf JE (1992) *J Phys Chem* 96: 9768
13. Shepard R (1993) *Theor Chim Acta* 84: 343
14. Hutter J, Parinello M, Vogel S (1994) *J Chem Phys* 101: 3862
15. Rendell AP (1994) *Chem Phys Lett* 229: 204
16. Wong AT, Harrison RJ (1995) *J Comput Chem* 16: 1291
17. Schmidt MW, Baldrige KK, Boats JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen, KA, Su S, Windus TL, Dupuis M, Montgomery JA (1993) *J Comput Chem* 14: 1347
18. Fletcher R (1980) *Practical methods of optimization*, vol. 1. Wiley, New York
19. Roothaan CCJ (1960) *Rev Mod Phys* 32: 179
20. Bobrowicz FW, Goddard III, WA (1977) In: Schaefer III, HF (ed) *Modern theoretical chemistry: methods of electronic structure theory*, vol. 3. Plenum, New York, p 79
21. Yaffe LG, Goddard III, WA (1976) *Phys Rev A* 13: 1682
22. Levy B, Berthier G (1969) *Int J Quantum Chem* 2: 307; erratum: *Int J Quantum Chem* (1969) 3: 247
23. Ruedenberg K, Cheung LM, Elbert ST (1979) *Int J Quantum Chem* 16: 1069
24. Roos BO (1980) *Int J Quantum Chem* S14: 175
25. Dalgaard E, Jørgensen P (1978) *J Chem Phys* 69: 3833
26. Siegbahn P, Heiberg A, Roos B, Levy B (1980) *Physica Scripta* 21: 323
27. Shepard R (1987) *Adv Chem Phys* 69: 63
28. Meier U, Staemmler V (1989) *Theor Chim Acta* 76: 95
29. Dupuis M, Chin S, Marquez A (1994) In: Malli GL (ed) *Relativistic and electron correlation effects in molecules and solids*. Plenum Press, New York, p 315
30. Camp N, Nicholas R, King HF (1982) *J Chem Phys* 77: 3056
31. Requests for the input files used in these examples can be addressed to M. W. Schmidt by e-mail: [mike@si.fi.ameslab.gov](mailto:mike@si.fi.ameslab.gov)