Theor Chim Acta (1993) 84: 343–351

Theoretica
Chimica Acta

Elimination of the diagonalization bottleneck in parallel Direct-SCF methods

Ron Shepard
Theoretical Chemistry Group, Chemistry Division, Argonne National Laboratory,
Argonne, IL 60439, USA

Received October 1, 1991/Accepted February 25, 1992

Summary. It is shown that the matrix diagonalization bottleneck associated with the *sequential* $O(\mathrm{N}^3_{\mathrm{BFN}})$ diagonalization of the fock matrix within each iteration of the Direct-SCF procedure may be eliminated, and replaced instead with a combination of *parallel* $O(\mathrm{N}^{<4}_{\mathrm{BFN}})$ and *sequential* $O(\mathrm{N}^3_{\mathrm{Sub}})$ steps. For large basis sets, the relation $\mathrm{N}_{\mathrm{Sub}} \ll \mathrm{N}_{\mathrm{BFN}}$ between the dimension of the expansion subspace and the number of basis functions leads to a method of wave-function optimization in which the sequential bottleneck is eliminated. As a side benefit, the second-order iterative procedure on which this method is based displays superior convergence properties, and provides greater insight into the behavior of the energy with respect to orbital variations, than the traditional first-order, fixed-point, iterative approaches. The implementation of this method may be incorporated into essentially any existing Direct-SCF program with only minimal, and localized, changes.

Key words: Diagonalization – Bottleneck – Parallel – Direct-SCF – Second-order convergence

1 Introduction

In traditional Self-Consistent-Field (SCF) wave-function optimization methods [1, 2], an effective one-electron hamiltonian operator, called the fock operator, is constructed within a given orbital basis. This fock operator depends on the (occupied) orbitals, or equivalently within the given orbital basis, on the expansion coefficients of these orbitals. This dependence will be denoted $\boldsymbol{F}[\boldsymbol{C}]$, where $\boldsymbol{C}$ is the matrix of orbital coefficients, $\boldsymbol{F}$ is the associated fock matrix, and the functional dependence is denoted with the square brackets. The orbital basis is usually called the *atomic orbital* (AO) basis, and the orthonormal orbitals which define the wave function are called the *molecular orbitals* (MO). In traditional approaches, a solution to the nonlinear equation:

$$\boldsymbol{F}[\boldsymbol{C}]\boldsymbol{C} = \boldsymbol{S}\boldsymbol{C}\boldsymbol{\varepsilon} \tag{1}$$

is sought using various fixed-point iterative schemes. In this equation, the matrix $\boldsymbol{S}$ is the overlap matrix of the basis functions, and serves as a metric for the orbital variations. The elements of the diagonal matrix $\boldsymbol{\varepsilon}$ are usually interpreted as the orbital energies, and along with other available information may be used to compute the total energy which is defined as the expectation value of the hamiltonian operator.

$$E = \langle\psi|H|\psi\rangle/\langle\psi|\psi\rangle \tag{2}$$

The usual iterative scheme may be denoted as:

$$\boldsymbol{F}[\boldsymbol{C}^{(i)}]\boldsymbol{C}^{(i+1)} = \boldsymbol{S}\boldsymbol{C}^{(i+1)}\boldsymbol{\varepsilon}^{(i+1)} \tag{3}$$

where the superscripts indicate a running iteration index. That is, given a set of occupied orbitals $\boldsymbol{C}^{(i)}$, the fock matrix is constructed. This fock matrix $\boldsymbol{F}[\boldsymbol{C}^{(i)}]$ is then used in a generalized eigenvalue equation to define a new set of orbitals for the next iteration $\boldsymbol{C}^{(i+1)}$. Convergence is determined usually by monitoring the change of the molecular orbitals between successive iterations, and stopping the procedure when the change is sufficiently small. Additional convergence criteria might include also changes in the total energy, in the one-particle density matrix in the AO basis $\boldsymbol{D}^{[\mathrm{AO}]}$, or in other properties of interest. Convergence difficulties are often encountered with the above iterative method [3], and consequently much effort has been devoted to the development of suitable damping and interpolation schemes [3, 4]. The notion of a self-consistent-field arises from Eq. (1) using the interpretation that the occupied orbitals define an *effective* or *averaged* electrostatic field, within which each electron moves independently of the other electrons.

In Direct-SCF approaches [5–8], the I/O of the electron repulsion integrals associated with the construction of the fock matrix within each iteration is eliminated, and replaced instead with the recomputation of these integrals as required. In sequential implementations, the greatest benefit of this approach is the elimination of the accompanying filespace requirements, thereby allowing computations with several hundred or even a few thousand basis functions to be completed on supercomputers and dedicated workstations. An active area of research involves the exploitation of the sparseness of the repulsion integral array. For large molecular systems, the effort for Direct-SCF calculations scales empirically only as $O(\sim \mathrm{N}_{\mathrm{BFN}}^{3})$, instead of the small-basis limit $O(\mathrm{N}_{\mathrm{BFN}}^{4})$, when this sparseness is used effectively with integral cutoff and Δ-density matrix thresholds. Herein, such dependence will be denoted $O(\mathrm{N}_{\mathrm{BFN}}^{<4})$. The overall result is that, for sufficiently large basis sets, Direct-SCF procedures are actually more efficient than conventional I/O-based procedures, despite the redundant repulsion integral recomputations.

Equation (1) follows from the variation of the total energy of Eq. (2), subject to the constraint that the orbitals remain orthonormal, $\boldsymbol{C}^{\dagger}\boldsymbol{S}\boldsymbol{C} = \mathbf{1}$. The derivation of Eq. (1) is straightforward, but somewhat lengthy, and will not be given in this brief report (see Ref. [2], for example, for further details). The theory and interpretation of electronic wave functions have been influenced strongly by the ideas of an effective one-electron hamiltonian operator, the general form of Eq. (1), the orbital energies defined by Eq. (1), and even the particular *canonical* orbitals which result from Eq. (1). Theories related to solid state and band structure calculations, the electron gas, diagrammatic perturbation theory, time-dependent perturbation theory, ionization potentials, electron affinities, electron propagators, and Green's functions, to name only a few, have all been influenced by the form of Eq. (1).

However, Eq. (1) actually results from three *independent* conditions: (1) that the molecular orbitals remain orthonormal, (2) that the total energy is stationary with respect to orbital variations, and (3) that the orbitals are chosen to be in their canonical form. In the more recently developed multiconfiguration SCF (MCSCF) theories in which the role of an effective one-electron hamiltonian operator is less important, these three conditions have been separated cleanly from each other. Orthonormalization is enforced by the particular choice of wave function variation coordinates. Canonicalization of the orbitals in MCSCF calculations is only applied as an afterthought, if at all, and even when it is applied, it is just as likely to be in the form of natural orbital resolution conditions (i.e. a diagonal density matrix constraint), or in the form of orbital localization conditions, as in the form of a diagonal fock matrix constraint (see, for example, Ref. [9] for a discussion of the orbital resolution effects on energy gradients). However, during the solution of the orbital optimization equations, orbital canonicalization plays essentially no role whatsoever. Instead, the energy expectation value of Eq. (2) is simply regarded as a function of all possible (essential) orbital variations, and the energy is minimized within this variational space [10]. In this report, it is emphasized that this alternative viewpoint may also be applied to the case of single-configuration SCF wave functions.

It is then rather a matter of taste as to which principle is considered to be the more basic to the nature of the SCF wave function: the minimization of the energy with respect to orbital variations (i.e. minimization on an energy surface), or the determination of a self-consistent effective one-electron hamiltonian operator (i.e. a fock operator). This issue, although perhaps important in another context, is not discused further in this report. Instead, it will be demonstrated that the *energy-surface* point of view leads directly to an iterative method which allows for the optimization of SCF wave functions without the need for a matrix diagonalization. This optimization method applies to both closed- and open-shell and to both spin-restricted (RHF) and spin-unrestricted (UHF) wave functions. The elimination of the matrix diagonalization step is particularly important in parallel implementations of the Direct-SCF method. This is due to the fact that matrix diagonalization is particularly difficult to implement efficiently on either distributed or shared memory machines, and on either single-instruction-multiple-data (SIMD) or multiple-instruction-multiple-data (MIMD) architectures, due to the necessary compromises between load balancing and the communication overhead. During the workshop, F. E. Harris and R. Littlefield [11] both discussed the difficulties associated with matrix diagonalization on parallel machines, and these presentations provided the impetus for the preparation of this report.

2 Parallel second-order SCF methods

Second-order SCF methods applicable to sequential implementation have been developed previously by Bacskay [12]. A brief summary of this approach is given here, primarily to establish notation and to point out the important features for a parallel implementation. Due to the similarities to the more general MCSCF case, it is most convenient to use a similar approach in the single-configuration SCF case. Consequently, the notation of Ref. [10] will be used and extended where necessary. The simple case of a closed-shell RHF wave function will be considered explicitly, but extension to more general cases [3, 12] is straightforward.

It is most convenient to begin the discussion by assuming that an orthonormal MO basis is available, and then to express the required computational steps in the AO basis in which Direct-SCF-type integral-recomputation technology may be applied. To this end, assume that the AO basis set χ and an orthonormal MO basis $\boldsymbol{\phi}$ are related by the orbital coefficient matrix $\boldsymbol{C}$ as:

$$\boldsymbol{\phi} = \chi \boldsymbol{C} \tag{4}$$

A transformation to a new orthonormal basis may be written as $\boldsymbol{\phi}' = \boldsymbol{\phi}\boldsymbol{U}$ where $\boldsymbol{U}$ is a real rotation matrix (i.e. $\boldsymbol{U}^T\boldsymbol{U} = \mathbf{1}$; $Det(\boldsymbol{U}) = 1$) parameterized as $\boldsymbol{U} = \exp(\boldsymbol{K})$ with $\boldsymbol{K} = -\boldsymbol{K}^T$. If all of the occupied orbitals are grouped together, then the *essential* orbital rotation parameters (i.e. nonredundant) occur only in the off-diagonal-block of the antisymmetric matrix $\boldsymbol{K}$. These elements will be written as $\boldsymbol{K}_{ia}$ ($= -\boldsymbol{K}_{ai}$) for i occupied and a unoccupied in the wave function. The energy does not depend on the other elements of the matrix, i.e. K_{ij} and K_{ab}, and they may be set to zero without loss of generality. For simplicity, the complications due to spatial symmetry will not be considered (see, for example, Ref. [10] for further details), but it is straightforward to enforce symmetry constraints on the matrices $\boldsymbol{K}$ and $\boldsymbol{U}$. With this parameterization of the wave-function variations, the energy may be written explicitly as an expansion about the current wave function in terms of the parameters $\{K_{ia}\}$.

$$E(\boldsymbol{k}) = E(0) + \boldsymbol{w}^T\boldsymbol{k} + \tfrac{1}{2}\boldsymbol{k}^T\boldsymbol{B}\boldsymbol{k} + O(\boldsymbol{K}^3)\ldots \tag{5}$$

with the elements of the vector $\boldsymbol{k}$ defined as $k_{(ia)} = K_{ia}$ with (ia) treated as a single vector index. The elements of the gradient vector $\boldsymbol{w}$ and the hessian matrix $\boldsymbol{B}$ are given as:

$$w_{(ia)} = 2F_{ia} \tag{6}$$

$$B_{(ia)(jb)} = 2F_{ab}\delta_{ij} - 2F_{ij}\delta_{ab} + 8(2g_{aibj} - \tfrac{1}{2}g_{ajbi} - \tfrac{1}{2}g_{abij}) \tag{7}$$

The fock matrix $\boldsymbol{F}$ in these expressions is defined in the MO basis as:

$$F_{uv} = 2h_{uv} + \sum_{j}^{occ} 4g_{uvjj} - 2g_{ujvj} \tag{8}$$

$$= 2h_{uv} + \sum_{ij}^{occ} (2g_{uvij} - \tfrac{1}{2}g_{uivj} - \tfrac{1}{2}g_{ujvi})2\delta_{ij} \tag{9}$$

The arrays $\boldsymbol{h}$ and $\boldsymbol{g}$ in these expressions are the usual one-electron hamiltonian and two-electron repulsion arrays respectively. The matrix $(2\delta_{ij}) = D_{ij}^{[\mathrm{MO}]}$ for i, j occupied is the one-electron density matrix in the MO basis. (Note that this definition of $\boldsymbol{F}$ differs by a factor of two from that usually used for SCF cases, but is rather more consistent with that used elsewhere in electronic structure theory.) The second form of the fock matrix $\boldsymbol{F}$ in Eq. (9) is most convenient for back-transformation to the AO basis.

$$F_{uv} = \sum_{\mu\lambda} C_{\mu u}C_{\lambda v}\left(2h_{\mu\lambda} + \sum_{\alpha\beta}(2g_{\alpha\beta\mu\lambda} - \tfrac{1}{2}g_{\alpha\mu\beta\lambda} - \tfrac{1}{2}g_{\alpha\lambda\beta\mu})D_{\alpha\beta}^{[\mathrm{AO}]}\right) \tag{10}$$

$$= (\boldsymbol{C}^T\boldsymbol{F}^{[\mathrm{AO}]}\boldsymbol{C})_{uv} \tag{11}$$

$$= (\boldsymbol{C}^T(2\boldsymbol{h}^{[\mathrm{AO}]}) + \boldsymbol{Q}[\boldsymbol{D}^{[\mathrm{AO}]}]\boldsymbol{C})_{uv} \tag{12}$$

In the last expression, the two-electron contributions have been separated for later reference. The matrix $\boldsymbol{Q}[\boldsymbol{D}^{[\mathrm{AO}]}]$ is the result of the two-index contraction in Eq. (10) of repulsion integrals $\boldsymbol{g}$ in the AO basis with the AO density matrix $\boldsymbol{D}^{[\mathrm{AO}]}$, defined as $\boldsymbol{D}^{[\mathrm{AO}]} = \boldsymbol{C}\boldsymbol{D}^{[\mathrm{MO}]}\boldsymbol{C}^T$. It is the $\boldsymbol{F}^{[\mathrm{AO}]}$ matrix, which corresponds to $\boldsymbol{F}[\boldsymbol{C}]$ in Eqs. (1) and (3), that is constructed in traditional SCF optimization methods.

Optimization of the parameters $\{K_{ia}\}$ based on truncation at second order of Eq. (5) leads directly to the Newton–Raphson procedure. This optimization method is seldom used in MCSCF optimizations due to its poor "global" behavior, and several alternatives have been developed. Bacskay [12] has discussed this issue in the SCF case, and additional discussion and references may be found in Refs. [10, 13] for the MCSCF case. It has been found sufficient in the general MCSCF case to formulate the determination of the correction vector in terms of subspace representations of the gradient vector $\boldsymbol{w}$ and hessian matrix $\boldsymbol{B}$. Herein we assume that this is the case for the simpler SCF wave functions. To this end, a set of linearly independent vectors $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots$ are collected together to form the columns of the matrix $\boldsymbol{X}$. The subspace representations of the gradient and hessian are then given by $\mathsf{w} = \boldsymbol{X}^T\boldsymbol{w}$ and $\mathsf{B} = \boldsymbol{X}^T\boldsymbol{B}\boldsymbol{X}$ respectively. The correction vector $\boldsymbol{k}$ will ultimately be represented as a linear combination of these trial vectors as $\boldsymbol{k} = \boldsymbol{X}\mathsf{k}$. The subspace vector k depends of course on the details of the iterative procedure; these details are unimportant to the present discussion. For definiteness, it will be assumed that k is the solution of the subspace representation of the rational functional approximation to Eq. (5). The subspace vector k is then determined from the solution to the eigenvector equation [10]:

$$\begin{pmatrix} \mathsf{B} & \mathsf{w} \\ \mathsf{w}^T & 0 \end{pmatrix} \begin{pmatrix} \mathsf{k} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathsf{k} \\ 1 \end{pmatrix} \lambda \tag{13}$$

Although this is an adequate choice based on previous MCSCF experience, particularly when combined with trust-radius and level-shifting options [10], it should be stressed that such a choice may not be optimal for the optimization of simpler SCF wave functions. In any case, the matrix vector products $\boldsymbol{\sigma} = (\boldsymbol{B}\boldsymbol{x})$ are required in order to compute the subspace matrix B, and we now focus on this task.

In the MO basis, the matrix-vector product may be written

$$\sigma_{(ia)} = \sum_{(jb)} B_{(ia)(jb)} x_{(jb)} \tag{14}$$

$$= 2\sum_b F_{ab} x_{ib} - 2\sum_j F_{ij} x_{ja} + 4\sum_{jb} (2g_{aibj} - \tfrac{1}{2}g_{ajbi} - \tfrac{1}{2}g_{abij})(2x_{jb}) \tag{15}$$

The first two terms are simple matrix products, and need not be discussed further other than to mention that efficient procedures for performing this common operation are available on essentially all parallel machines. Back transformation of the last contribution to the AO basis results in the final expression:

$$\boldsymbol{\sigma} = 2(\boldsymbol{x}\boldsymbol{F}) - 2(\boldsymbol{F}\boldsymbol{x}) + 4\boldsymbol{C}^T\boldsymbol{Q}[\boldsymbol{Z}^{[\mathrm{AO}]}]\boldsymbol{C} \tag{16}$$

with the intermediate symmetric matrix $\boldsymbol{Z}^{[\mathrm{AO}]}$ defined as:

$$Y_{\mu\nu} = \sum_{jb} C_{\mu j} x_{jb} C_{\nu b} = (\boldsymbol{C}\boldsymbol{x}\boldsymbol{C}^T)_{\mu\nu} \tag{17}$$

$$\boldsymbol{Z}^{[\mathrm{AO}]} = \boldsymbol{Y} + \boldsymbol{Y}^T \tag{18}$$

(Note that $\boldsymbol{\sigma}$ and $\boldsymbol{x}$ are treated as matrices in Eqs. (15–17), with index association $x_{(ia)} \rightarrow x_{ia}$.) As seen in Eq. (16), the two-electron repulsion integral contributions to the required matrix-vector product may be computed in exactly the same manner as the construction of the $\boldsymbol{F}^{[\mathrm{AO}]}$ matrix of Eq. (11). The only difference is that the intermediate matrix $\boldsymbol{Z}^{[\mathrm{AO}]}$ is used in place of the density matrix elements $\boldsymbol{D}^{[\mathrm{AO}]}$.

Once the orbital correction matrix $\boldsymbol{K}$ has been determined using some iterative procedure, there remains the task of computing the transformation matrix $\boldsymbol{U} = \exp(\boldsymbol{K})$. In usual MCSCF approaches, this is done by factoring the antisymmetric $\boldsymbol{K}$ as $\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T$ and forming $\boldsymbol{U} = \boldsymbol{V}\exp(\boldsymbol{\Delta})\boldsymbol{V}^T$ where $\boldsymbol{V}$ is orthogonal and $\boldsymbol{\Delta}$

is block diagonal and contains at most 2 × 2 subblocks. This is essentially equivalent in effort and complexity to diagonalization of a symmetric matrix, and therefore this approach cannot be used efficiently in a parallel implementation. The author is not aware of any practical alternative method for the exact computation of $\boldsymbol{U}$ that avoids a comparable diagonalization step [14]. However, there are two reasonable alternatives to the exact computation of $\boldsymbol{U}$. The first is based on the truncated expansion:

$$\boldsymbol{U} = \mathbf{1} + \boldsymbol{K} + \tfrac{1}{2}\boldsymbol{K}^2 \qquad (19)$$

followed by a Schmidt orthonormalization step. The expansion could be continued further, but second order is sufficient to guarantee overall second order convergence of the iterative procedure in the neighborhood of the final result. This orthonormalization procedure, although consuming roughly the same arithmetic operations as diagonalization $O(\mathrm{N}_{\mathrm{BFN}}^3)$, does not suffer the same communication and load-balancing bottleneck as matrix diagonalization.

The second procedure involves the construction of a rational approximation to the matrix exponential. One such approximation, accurate to second order, is:

$$\boldsymbol{U} = (\mathbf{1} - \tfrac{1}{2}\boldsymbol{K})^{-1}(\mathbf{1} + \tfrac{1}{2}\boldsymbol{K}) \qquad (20)$$

It is easily verified that $\boldsymbol{U}$ is orthogonal, so no additional orthonormalization step is required. Generally it would be expected that this method would be the best overall approach, since essentially all interesting architectures are designed to solve linear equations at their most impressive computational rates. For example, the latest version of the CM-2 Connection Machine (a SIMD architecture) achieves 9.03 GFLOPS double precision and 17 GFLOPS single precision on the large-scale LINPACK benchmark [15]. The Intel Touchstone Delta machine (a MIMD architecture) achieves 11.9 GFLOPS on the double precision LINPACK benchmark [16]. This popular linear algebra benchmark is based on essentially the same computational step as that required in Eq. (20).

With these computational steps discussed above, the entire SCF iterative procedure may now be outlined.

```
Initialize $\boldsymbol{C}$
Compute $D_{\mu\nu}^{[\mathrm{AO}]} = 2\sum_{i}^{occ} C_{\mu i}C_{\nu i}$
DO                                                      !SCF Iteration
   Compute $\boldsymbol{F}^{[\mathrm{AO}]}$                                      !Parallel $N_{BFN}^{<4}$ work
   Compute $\boldsymbol{F}^{[\mathrm{MO}]}$                                      !Parallel $N_{BFN}^{3}$ work
   Initialize $\boldsymbol{x}^1$
   DO N = 1 TO NMAX                                     !Subspace iterations
      Compute $\boldsymbol{Z}^{[\mathrm{AO}]}$ from $\boldsymbol{x}^N$ using Eq. (18)            !Parallel $N_{BFN}^{3}$ work
      Compute $\boldsymbol{Q}[\boldsymbol{Z}^{[\mathrm{AO}]}]$ using Eq. (16)               !Parallel $N_{BFN}^{<4}$ work
      Compute $\boldsymbol{\sigma}^{\mathrm{N}}$ using Eq. (16)                     !Parallel $N_{BFN}^{3}$ work
      Compute $\mathsf{w}$, $\mathsf{B}$, $\mathsf{k}$ and form $\boldsymbol{k}$                  !Sequential $N_{Sub}^{3}$ work
      IF (CONVERGED) THEN EXIT
      Construct $\boldsymbol{x}^{N+1}$
   ENDDO
   Construct $\boldsymbol{K}$ from the final $\boldsymbol{k} = \boldsymbol{X}\mathbf{k}$
   Construct $\boldsymbol{U}$ from Eq. (20)                          !Parallel $N_{BFN}^{3}$ work
   Update $\boldsymbol{C} \leftarrow \boldsymbol{CU}$                                 !Parallel $N_{BFN}^{3}$ work
ENDDO
```

As seen in this outline, the only sequential step of any possible significance involves the subspace solution for **k** and formation of $\boldsymbol{k}$. The effort required for this step varies, depending on the iterative algorithm. The N_{Sub}^3 effort given in the outline is that required for the eigenvector solution of Eq. (13), but this should be representative also of other iterative subspace methods.

In the usual MCSCF scheme, the subspace is initialized with one vector, usually taken as the gradient vector for that iteration, $x^1_{(ia)} = w_{(ia)} = 2F^{[MO]}_{ia}$. Each subspace iteration then adds a new expansion vector, so that the dimension of the subspace grows by one each iteration. However in the SCF case, due to the fact that the parallel $N_{\text{BFN}}^{\leq 4}$ effort each subspace iteration is largely independent of the number of new trial vectors, other possibilities would appear more appropriate. Assume, for example, that the computation of each repulsion integral on average requires about 100 arithmetic operations. Multiplication of this integral by the corresponding $\boldsymbol{Z}^{[AO]}$ elements consumes only $\sim$6 additional arithmetic operations, depending on the number of coincident orbital indices. Consequently, the computation of several $\boldsymbol{Q}$ matrices simultaneously (say 10–20 at a time) might require only an insignificant additional effort compared to the computation of a single $\boldsymbol{Q}$ matrix. In this case, the subspace dimension would increase by this larger number for that iteration, ignoring any reductions resulting from linear-dependence of the correction vectors near convergence, and the total number of subspace iterations would be reduced due to the increased variational flexibility. Reductions from 4–6 subspace cycles to 1–3 cycles might be expected in practice. One possible way to generate such vectors initially would be to use correction vectors generated during the previous SCF iterations. These vectors, roughly speaking, point back from the current wave function toward the wave functions of the previous iterations, automatically providing flexibility in the most important directions of wave function variational space, and avoiding to some extent oscillatory behavior within the subspace iterations. It is straightforward, of course, to reduce the subspace dimension at any time by taking the most appropriate linear combinations and resetting the subspace dimension counter. For example, the lowest few eigenvectors from the subspace eigenvalue equation of Eq. (13) could be used to define this transformation. This approach has worked successfully in the more general MCSCF case. This subspace dimension reduction ensures that the effort required for the sequential subspace manipulations never becomes prohibitive.

One advantage of the proposed subspace approach is that the accuracy of the final solution does not depend directly on the precision of the elements of the individual expansion vectors, provided the matrix-vector products are computed correctly for the given expansion vectors. It has been shown previously that such subspace expansion methods lend themselves naturally to the use of rather severe truncations in the individual elements of the expansion vectors [17]. In the present context, the individual expansion vectors may be truncated, normalized to small integer values, and then subjected to further data-compression techniques. Such data compression is particularly important in distributed-memory parallel computing environments where the available local memory per node is often scarce and the communication overhead is relatively large. Such data-compression techniques cannot be applied directly to traditional fixed-point iterative schemes since the truncated density matrix loses idempotency, resulting in an inaccurate fock matrix.

As discussed previously, the converged orbitals from the above procedure generally do not result in a diagonal fock matrix $\boldsymbol{F}^{[MO]}$. However, if desired, such a canonicalization may be imposed after the optimization procedure has completed. This canonicalization, if restricted to rotations among occupied orbitals only and/or to rotations among unoccupied orbitals only, does not affect the wave function, but only the *representation* of the wave function. Since this canonicalization does not affect the wave function, it does not affect any observable property of the wave function. The energy, electron density, dipole moment, etc. are all invariant to such redundant orbital rotations. Apart from any implicit iterations within the diagonalization procedure itself, this final canonicalization is a noniterative process. That is, the $\boldsymbol{C}^{(i+1)}$ matrix of Eq. (3) resulting from the final $\boldsymbol{F}[\boldsymbol{C}^{(i)}]$ matrix produces a $\boldsymbol{F}[\boldsymbol{C}^{(i+1)}]$ matrix that satisfies Eq. (1) to numerical accuracy; no further iterations are required to achieve self-consistency.

3 Conclusions

A practical method of SCF wave function optimization has been presented based on the idea of multidimensional variation of the energy expectation value. This method eliminates the matrix diagonalization bottleneck associated with traditional fock-operator-based SCF iterative procedures. These traditional approaches attempt to simultaneously optimize and canonicalize the occupied orbitals. The proposed method retains the idea of optimization, but ignores the superfluous canonicalization constraint. Since the proposed procedure is based on a well-understood second-order iterative algorithm, improved convergence, compared to the usual fixed-point iterative procedures will be observed. In order to improve efficiency beyond that expected for existing sequential implementations of second-order methods for SCF and MCSCF wave function optimization, several suggestions for further investigations are given. These include the idea of computing several matrix-vector products simultaneously, the reuse of vectors from previous iterations in order to span the most important regions of wave function variation space with the least amount of overall effort, and the use of data compression techniques to optimize the storage and communication of the subspace expansion vectors.

Acknowledgements. The author acknowledges informative discussions concerning various aspects of this work with several of the workshop participants including Professors J. Almlöf, F. E. Harris, and H. F. King, and Drs. T. H. Dunning, Jr., M. Feyereisen, R. J. Harrison, and R. A. Kendall. This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, under contract W-31-109-ENG-38.

References

1. Roothaan CCJ (1951) Rev Mod Phys 23:69
2. Szabo A, Ostlund NS (1989) Modern quantum chemistry. McGraw-Hill, NY
3. Schlegel HB, McDouall JJW (1991) in: Ögretir C, Csizmadia I (eds) Computational advances in organic chemistry: Molecular structure and reactivity. NATO ASI Series C 330. Kluwer, Dordrecht, p 167
4. Pulay P (1980) Chem Phys Lett 73:393; Pulay P (1982) J Computational Chem 3:556
5. Almlöf J, Faegri K, Korsell K (1982) J Computational Chem 3:385
6. Price SL, Harrison RJ, Guest MF (1989) J Computational Chem 10:552
7. Häser M, Ahlrichs R (1989) J Computational Chem 10:104

8. GAUSSIAN 88: Frisch MJ, Head-Gordon M, Schlegel HB, Raghavachari K, Binkley JS, Gonzalez C, Defrees DJ, Fox DJ, Whiteside RA, Seeger R, Melius CF, Baker J, Martin R, Kahn LR, Stewart JJP, Fluder EM, Topiol S, Pople JA. Gaussian Inc., Pittsburgh, PA, 1988
9. Shepard R, Shavitt I, Pitzer RM, Comeau DC, Pepper M, Lischka H, Szalay PG, Ahlrichs R, Brown FB, Zhao JG (1988) Int J Quantum Chem S22:149; Shepard R, Lischka H, Szalay PG, Kovar T, Ernzerhof M (1992) J Chem Phys 96:2085
10. Shepard R (1987) in: Lawley KP (ed) *Ab initio* methods in quantum chemistry II, Advances in Chemical Physics 69. Wiley, NY, p 63
11. Littlefield R (this issue)
12. Bacskay GB (1981) Chem Phys 61:385
13. Shepard R (1992) Theor Chim Acta 84:55–83
14. Moler C, van Loan C (1978) SIAM Rev 20:801
15. Browne T (1991) public communication via the *comp.parallel* newsgroup, June, 1991. Tim Browne may be contacted via email at browne@think.com.
16. Baugh J (1991) public communication via the *comp.parallel* newsgroup, August, 1991. Jerry Baugh may be contacted via email at baugh@ssd.intel.com. Intel Touchstone benchmarks are made available also by R. A. van de Geijn at the email address rvdg@cs.utexas.edu.
17. Shepard R (1990) J Computational Chem 11:45