TABLE II: Hamiltonian Matrix,^{a} B_1 Overall Symmetry

	(A_1B_1)	(B_1A_1)	(B_2B_3)	(B_3B_2)
(A_1B_1)	ER			
(B_1A_1)	**SO$_z$**	ER		
(B_2B_3)	SO$_y$	SO$_x$	ER	
(B_3B_2)	SO$_x$	SO$_y$	**SO$_z$**	ER

^{a} (Spatial symmetry,spin symmetry); ER denotes electron repulsion matrix element; SO$_x$ denotes xx spin–orbit matrix element, etc.; pure imaginary matrix elements are boldfaced; all others are real.

siderations only of D_{2h}' and its subgroups.

Reality of the Hamiltonian Matrix

The most computationally convenient general basis functions are real atomic and molecular orbitals and the α,β spin functions. In these bases, the matrix elements of $\hat{H}^{SO}$ are, in general, complex, and therefore the wave functions (configuration-interaction coefficients) are also complex. When the point-group symmetry of the system is C_{2v}, D_2, D_{2h}, or higher, however, it is possible to show that the Hamiltonian matrix elements are either pure real or pure imaginary and that a simple redefinition of the many-electron functions can make the entire matrix real. The resulting computational savings in memory and in diagonalization time are quite important.

The integrals of $\hat{l}$ and $\hat{s}$ in the real-orbital basis and α,β basis are either pure real or pure imaginary:

pure real: $\qquad\qquad \hat{s}_x, \hat{s}_z$

pure imaginary: $\qquad \hat{l}_x, \hat{l}_y, \hat{l}_z, \hat{s}_y$

Accordingly, integrals of $\hat{l}_x\hat{s}_x$ and $\hat{l}_z\hat{s}_z$ are pure imaginary, and those of $\hat{l}_y\hat{s}_y$ are pure real. When the system has D_2' symmetry, the x, y, and z components of the angular momentum operators each transform according to different irreducible representations (Table I), so that matrix elements of $\hat{H}^{SO}$ between symmetry-adapted functions can contain a nonzero contribution from at most one of the x, y, or z terms in the $\hat{l}\cdot\hat{s}$ dot product. The matrix elements

(10) Shavitt, I. *Chem. Phys. Lett.* **1979**, *63*, 421.

TABLE III: Phase Factors^a To Obtain Real Hamiltonian Matrices

$(A_1A_1\|A_1)$	1	$(A_1B_2\|B_2)$	1	$(A_1E_+\|E_+)$	1
$(B_1B_1\|A_1)$	i	$(B_1B_3\|B_2)$	i	$(B_1E_+\|E_+)$	i
$(B_2B_2\|A_1)$	1	$(B_2A_1\|B_2)$	1	$(B_2E_-\|E_+)$	1
$(B_3B_3\|A_1)$	i	$(B_3B_1\|B_2)$	i	$(B_3E_-\|E_+)$	i
$(A_1B_1\|B_1)$	1	$(A_1B_3\|B_3)$	1	$(A_1E_-\|E_-)$	1
$(B_1A_1\|B_1)$	i	$(B_1B_2\|B_3)$	i	$(B_1E_-\|E_-)$	$-i$
$(B_2B_3\|B_1)$	1	$(B_2B_1\|B_3)$	1	$(B_2E_+\|E_-)$	-1
$(B_3B_2\|B_1)$	i	$(B_3A_1\|B_3)$	i	$(B_3E_+\|E_-)$	i

^{a} For D_2' group in the order (spatial,spin|overall); factors for C_{2v}' and D_{2h}' follow directly.

must therefore be either pure real or pure imaginary, as in the example in Table II, which shows a schematic lower half Hamiltonian matrix for wave functions of B_1 overall symmetry and the four possible pairs of (spatial symmetry, spin symmetry).

Inspection of Table II shows that the simple insertion of a factor of i into the definition of the (B_1A_1) and (B_3B_2) many-electron basis functions will make the matrix real everywhere and therefore symmetric instead of (complex) Hermitian, with obvious computational benefits. These simple symmetry and matrix-element considerations thus identify the imaginary configuration-interaction coefficients in advance.

A set of factors of ± 1, $\pm i$ are given for all D_2' cases in Table III. The factors are given in coupling constant notation and are not unique. The choices made for E overall symmetry are such that the E_+ and E_- submatrices will be identical.

Summary

Several methods of simplifying and therefore extending the range of ab initio approximate relativistic electronic-structure calculations have been derived. Since perturbation theory is not used, they are applicable to molecules containing any atom regardless of position in the periodic table. Some spatial symmetry is needed to use all of the methods, but, for example, all linear molecules have more than enough symmetry. Computer programs using these methods have been tested and are currently being applied, with small configuration-interaction wave functions, to the potential curve crossings and avoided crossings of CuF and to the electronic states of uranocene.

Optimization of Wave Function and Geometry in the Finite Basis Hartree–Fock Method

Martin Head-Gordon and John A. Pople*

Department of Chemistry, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213
(Received: July 14, 1987; In Final Form: October 14, 1987)

In the finite basis Hartree–Fock (HF) method, the energy is a function of independent wave function and geometric variables, which are variational parameters. A definition of the wave function variables as a set of plane rotation angles is adopted, and partial derivatives of the energy with respect to these variables are obtained, leading to a new method for HF wave function determination. Partial derivatives of the energy with respect to geometric variables, with no assumption of wave function optimality, are also derived. Together these two sets of first derivatives permit the simultaneous optimization of geometry and wave function in the HF method. Examples of the application of this procedure to water and methanol are given.

1. Introduction

The finite basis Hartree–Fock (HF) method was introduced by Roothaan[1] and Hall[2] for closed-shell molecules and generalized to open-shell systems by Pople and Nesbet.[3] It is the standard procedure of ab initio molecular orbital (MO) theory and yields an approximate energy and wave function at a specified nuclear geometry. The addition of analytic first derivatives of the HF

energy with respect to nuclear coordinates, which was pioneered by Pulay in his "force method",[4,5] has made the calculation of equilibrium HF geometries for medium-sized molecules a routine procedure.[6,7]

(1) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.
(2) Hall, G. G. *Proc. R. Soc. London, A* **1951**, *205*, 541.
(3) Pople, J. A.; Nesbet, R. K. *J. Chem. Phys.* **1954**, *22*, 571.
(4) Pulay, P. *Mol. Phys.* **1969**, *17*, 197.
(5) Pulay, P. In *Modern Theoretical Chemistry*; Schaefer, H. F., Ed.; Plenum: New York, 1977; Vol. 4, pp 153–185.

*Author to whom correspondence should be addressed.

The Roothaan or self-consistent field (SCF) procedure for solving the HF equations at a given molecular geometry is basically[8]

(1) Specify a geometry, an atomic orbital (AO) basis set consisting of N functions, and an initial guess for the unknown MO coefficients.

(2) Calculate N^2 one-electron integrals and approximately $N^4/8$ two-electron integrals, and save them on a mass-storage device.

(3) Use the current MO coefficients and stored integrals to form the matrix of the HF Hamiltonian in the AO basis (the Fock matrix).

(4) Transform the Fock matrix into the basis of the current MO's and diagonalize it. The eigenvectors yield the new MO coefficients, and if they are significantly different from the old values, go back to (3).

The first difficulty with this procedure is that convergence problems are encountered quite frequently in the iterative step. Generally they may be overcome by using various extrapolation schemes such as those of Pulay[9] or by resorting to algorithms[10-13] other than the Roothaan procedure, often based on variables other than the MO coefficients. The MO coefficients are not independent variables due to the constraint imposed in HF theory forcing the MO's to be orthonormal.[8]

A second difficulty is that as larger and larger HF calculations are performed, storing the two-electron integrals becomes increasingly problematic due to their quartic dependence on basis set size. Additionally, it appears that technological development of the computing capabilities necessary to calculate the integrals has been outstripping development of the input/output capabilities necessary to handle the stored integrals. To address these problems, Almlöf[14] developed the "direct SCF" approach in which the two-electron integrals are recalculated in each iteration (step 3) and discarded as they are used. As the remaining quantities are only of size N^2, this has permitted very large HF calculations with over 500 basis functions.[14,15] However, the extra work involved in recalculating the integrals means that much more computer time is required relative to the traditional procedure.

The force method[5] for determining HF equilibrium geometries can be viewed as follows:

(1) Specify an initial molecular geometry.

(2) Perform a HF calculation at the current geometry, by the methods discussed above.

(3) Use the resulting MO coefficients to calculate the gradient of the energy with respect to the geometric parameters, $\mathbf{r}$.

(4) Use the gradient in conjunction with an optimization algorithm[16,17] to alter the geometry. If the gradient and/or change in geometry is still sufficiently nonzero, go back to (2).

The energy E is a function of geometric variables $\mathbf{r}$, and wave function variables θ. From a strictly mathematical point of view, minimizing $E(\mathbf{r},\theta)$ by separately varying $\mathbf{r}$ is step 4, subject to keeping θ optimum in step 2, is a rather unnatural procedure relative to minimizing E with respect to all variables at once.

However, its utility arose because in the Roothaan procedure, the number of geometry changes, and hence the number of time-consuming two-electron integral recalculations and gradient evaluations, are minimized by keeping the wave function optimized at all times. In the context of the direct SCF approach though, where the two-electron integrals must be recalculated whenever θ is changed, it is clearly undesirable to fully optimize the wave function at each trial geometry. A better procedure might be to instead vary all parameters simultaneously to obtain the equilibrium geometry and wave function.

Accordingly, in this paper we extend the direct SCF approach to simultaneous optimization of geometry and wave function. To treat wave function variables on the same footing as geometric variables, it is desirable to reformulate the HF equations in terms of independent variables θ rather than the MO coefficients. At a given geometry this yields an alternative to the Roothaan procedure based on minimizing E using derivatives of the energy with respect to θ. The addition of partial derivatives with respect to geometry allows the energy to be minimized with respect to $\mathbf{r}$ and θ simultaneously.

An important issue is whether a first-order (first derivative based) or a second-order (first and second derivative based) optimization scheme should be used. While the convergence characteristics of second-order schemes are generally much better, the extra work of obtaining and handling second derivatives may still make them less economical. In particular, existing second-order methods for HF wave function optimization[10,11] are costlier than simpler extrapolation schemes,[9] despite their quadratic convergence. Furthermore, computing second derivatives of the energy with respect to geometric variables is much costlier than first derivatives, principally due to obtaining second derivatives of the two-electron integrals. Therefore, a first-order scheme is likely to be most efficient for simultaneous optimization of wave function and geometry and is the approach we shall pursue here.

The remainder of this paper proceeds as follows. In section 2, we choose a definition for θ and obtain expressions for $\partial E/\partial \theta$ and $\partial E/\partial \mathbf{r}$. With these partial derivatives, the energy can be minimized with respect to all parameters simultaneously using standard optimization algorithms; in section 3 the implementation of this scheme is discussed. Finally, in section 4, we report some preliminary applications of the method and compare the results with existing techniques.

2. Theory

In this section, the problem of minimizing the energy in the finite basis HF method, subject to maintaining MO orthonormality, is addressed. The notation used is fairly standard. Expressions are given in terms of spin orbitals,[8] which can be simplified to the spatial orbital results actually implemented by integration over the spin coordinate. AO basis functions are denoted by subscripts μ, ν, λ, σ, ..., occupied spin orbitals by i, j, k, l, ..., unoccupied spin orbitals by a, b, c, d, ..., general spin orbitals by p, q, r, s, ..., eigenvectors of the AO overlap matrix by α, β, ..., and orthonormalized basis functions (OBF) by w, x, y, and z. For simplicity, the basis functions and spin orbitals are assumed to be real. Boldface upper case is used for matrices, while boldface lower case will indicate a column vector or a diagonal square matrix according to context.

The energy associated with a single determinant wave function follows from standard matrix element rules[8]

$$E = \sum_i H_{ii} + \tfrac{1}{2}\sum_{ij}(ij\|ij) + V_{nuc} \qquad (1)$$

where H_{ii} is the diagonal matrix element of the one-electron core Hamiltonian involving spin orbital χ_i, V_{nuc} is the Coulombic repulsion of the point nuclei, and the remaining term describes interactions between electrons, where

$$(pq\|rs) =$$
$$\int\int \chi_p(1)\chi_q(2)(1/r_{12})[\chi_r(1)\chi_s(2) - \chi_s(1)\chi_r(2)]\ d\tau_1\ d\tau_2$$

The occupied spin orbitals χ_i are expressed as a linear combination of AO's ω_μ

(6) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab-Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(7) Carsky, P.; Urban, M. *Ab-Initio Calculations*; Springer-Verlag: Berlin, 1980.

(8) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; MacMillan: New York, 1982.

(9) (a) Pulay, P. *Chem. Phys. Lett.* **1980**, *73*, 393. (b) Pulay, P. *J. Comput. Chem.* **1982**, *3*, 556.

(10) Douady, J.; Ellinger, Y.; Subra, R.; Levy, B. *J. Chem. Phys.* **1980**, *72*, 1452.

(11) Bacskay, G. B. *Chem. Phys.* **1981**, *61*, 385.

(12) (a) Fernández Rico, J.; Garcia de la Vega, J. M.; Fernández Alonso, J. I.; Fantucci, P. *J. Comput. Chem.* **1983**, *4*, 33. (b) Fernández Rico, J.; Paniagua, M.; Fernández Alonso, J. I.; Fantucci, P. *J. Comput. Chem.* **1983**, *4*, 41.

(13) (a) Carbó, R.; Domingo, Ll.; Peris, J. *Adv. Quantum Chem.* **1982**, *15*, 215. (b) Carbó, R.; Domingo, Ll.; Peris, J.; Novoa, J. *J. Mol. Struct. (Theochem)* **1983**, *93*, 15.

(14) Almlöf, J.; Faegri, K., Jr.; Korsell, K. *J. Comput. Chem.* **1982**, *3*, 385.

(15) Cremer, D.; Gauss, J. *J. Comput. Chem.* **1986**, *7*, 274.

(16) Schlegel, H. B. *Adv. Chem. Phys.* **1987**, *67*, 249.

(17) Fletcher, R. *Practical Methods of Optimization*; Wiley: New York, 1980; Vol. 1.

$$\chi_i = \sum_\mu C_{\mu i}\omega_\mu \tag{2}$$

where $\mathbf{C}$ is the MO coefficient matrix. The energy is then

$$E = \sum_{i\mu\nu} C_{\mu i}C_{\nu i}H_{\mu\nu} + \tfrac{1}{2}\sum_{ij\mu\nu\lambda\sigma} C_{\mu i}C_{\nu i}C_{\lambda j}C_{\sigma j}(\mu\lambda||\nu\sigma) + V_{\text{nuc}} \tag{3}$$

The spin orbitals are required to be orthonormal

$$\int \chi_i\chi_j\,\mathrm{d}\tau = \sum_{\mu\nu} C_{\mu i}S_{\mu\nu}C_{\nu j} = \delta_{ij} \tag{4}$$

where $\mathbf{S}$ is the AO overlap matrix and δ_{ij} is the Kronecker delta. The problem then is to minimize the energy (3), while satisfying the constraints (4).

The Roothaan method is based on using Lagrange multipliers[8] to incorporate the constraints into an augmented functional which can then be minimized in an unconstrained fashion to yield

$$\sum_\nu F_{\mu\nu}C_{\nu i} - \sum_{\nu j}S_{\mu\nu}C_{\nu j}\epsilon_{ji} = 0 \tag{5}$$

where ϵ is a matrix of Lagrange multipliers and $\mathbf{F}$ is the Fock matrix which is defined as

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda\sigma j}C_{\lambda j}C_{\sigma j}(\mu\lambda||\nu\sigma) \tag{6}$$

Additionally, the energy is invariant to unitary transformations of the occupied spin orbitals among themselves. If we choose the unitary transformation to diagonalize the matrix of Lagrange multipliers ϵ, then (5) becomes

$$\sum_\nu F_{\mu\nu}C_{\nu i} = \sum_\nu S_{\mu\nu}C_{\nu i}e_i \tag{7}$$

which are the Roothaan equations. As the Fock matrix (6) is itself a function of the MO coefficients, the Roothaan equations must be solved by the kind of iterative algorithm indicated in the introduction, to obtain O occupied MO's and V ($=N-O$) unoccupied or virtual MO's. Since only rotations that mix occupied and unoccupied orbitals affect the energy, the number of degrees of freedom that determine the energy is OV.

We now wish to construct a wave function of OV variables θ, which automatically satisfies (4), thereby eliminating the need to include Lagrange multipliers in our formulation. Unitary transformations preserve orthonormality, which makes them an ideal way to incorporate the constraints (4). Suppose then that the AO basis is orthonormalized by a transformation $\mathbf{M}(\mathbf{r})$ which depends on the nuclear coordinates

$$\sum_{\mu\nu} M_{\mu w}S_{\mu\nu}M_{\nu x} = \delta_{wx} \tag{8}$$

where $\mathbf{M}$ will in practice be a set of initial guess MO coefficients. The MO coefficients are obtained as a unitary transformation $\mathbf{U}(\theta)$ of these orthogonal basis functions (OBF's)

$$C_{\mu p}(\mathbf{r},\theta) = \sum_w M_{\mu w}(\mathbf{r})\,U_{wp}(\theta) \tag{9}$$

where any unitary transformation $\mathbf{U}$ by definition satisfies

$$\sum_r U_{rp}U_{rq} = \delta_{pq} \tag{10}$$

By using (8)–(10), the MO's will themselves be orthonormal. Hence (9) is a way of defining the dependence of the MO coefficients on the independent variables $(\mathbf{r},\theta)$, which incorporates the orthonormality constraints.

In choosing a definition of the OV independent wave function variables, our main requirement is that it should be possible to obtain partial derivatives of the energy without any restrictions on the values of the variables. This permits the use of efficient quasi-Newton optimization methods and leaves two principal alternatives for parameterizing the unitary transformation, first as a direct product of elementary plane rotations, and second as the exponential of an antisymmetric matrix. For reasons discussed in Appendix 1, where the exponential alternative is investigated, we use the direct product of OV plane rotations (Givens, or Jacobi transformations) which mix occupied and unoccupied OBF's

$$\mathbf{U} = \prod_i^O (\prod_a^V \mathbf{G}(\theta_{ai})) = \prod_{(ai)}^{(OV)} \mathbf{G}(\theta_{ai}) \tag{11}$$

The Givens transformation $\mathbf{G}(\theta_{ai})$ mixing OBF's i and a differs from the unit matrix only in rows and columns i and a

$$\mathbf{G}(\theta_{ai}) = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & \cos\theta_{ai} & & & \sin\theta_{ai} & & \\ & & & & 1 & & & & \\ & & & & & \ddots & & & \\ & & & & & & 1 & & \\ & & & -\sin\theta_{ai} & & & \cos\theta_{ai} & & \\ & & & & & & & 1 & \\ & & & & & & & & \ddots \\ & & & & & & & & & 1 \end{bmatrix} \tag{12}$$

Note that, since individual $\mathbf{G}(\theta_{ai})$ do not commute, a consistent ordering of the Givens transformations, such as indicated in (11), must be used. The general properties of this form of $\mathbf{U}$ have been studied,[18] and it has been applied to minimize the energy in the semiempirical PPP method at a fixed geometry.[19] In ab initio HF theory, it has been used by Carbo et al.,[13] who obtained the derivative of the energy with respect to a single rotation angle subject to all other angles being zero. We shall now use this definition of the independent variables θ to obtain general expressions for $\partial E/\partial\theta$, without restrictions on the values of θ.

To differentiate the energy with respect to θ, the chain rule must be used

$$\frac{\partial E}{\partial\theta_{ai}} = \sum_{wj}\frac{\partial E}{\partial U_{wj}}\frac{\partial U_{wj}}{\partial\theta_{ai}} \tag{13}$$

Differentiating the energy expression (3) with respect to U_{wj}, after substituting (9) for the MO coefficients, yields an expression for the first term of (13)

$$\frac{\partial E}{\partial U_{wj}} = \sum_{\mu\nu}2M_{\mu w}F_{\mu\nu}C_{\nu j} \tag{14}$$

which, like the Roothaan equations (7), is in terms of the Fock matrix (6). Derivatives of the elements of $\mathbf{U}$ with respect to θ_{ai} follow from the defining relations (11) and (12), noting that only one of the Givens transformations depends on a particular θ_{ai}

$$\frac{\partial \mathbf{U}}{\partial\theta_{ai}} = (\prod_{(bj)<(ai)} \mathbf{G}(\theta_{bj}))\frac{\partial \mathbf{G}(\theta_{ai})}{\partial\theta_{ai}}(\prod_{(bj)>(ai)} \mathbf{G}(\theta_{bj})) \tag{15}$$

The derivative of $\mathbf{G}(\theta_{ai})$ with respect to θ_{ai} has only four nonzero elements, which are at the intersection of rows and columns i and a

$$\frac{\partial \mathbf{G}(\theta_{ai})}{\partial\theta_{ai}} = \begin{bmatrix} 0 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 0 & & & & & & \\ & & & -\sin\theta_{ai} & & & \cos\theta_{ai} & & \\ & & & & 0 & & & & \\ & & & & & \ddots & & & \\ & & & & & & 0 & & \\ & & & -\cos\theta_{ai} & & & -\sin\theta_{ai} & & \\ & & & & & & & 0 & \\ & & & & & & & & \ddots \\ & & & & & & & & & 0 \end{bmatrix} \tag{16}$$

Equation 13, together with (14)–(16), comprises a tractable expression for $\partial E/\partial\theta$, as will be discussed in section 3.

Finding θ to satisfy $\partial E/\partial\theta = 0$ will only yield the same energy as the Roothaan procedure if varying θ allows all energetically nonredundant redundant sets of MO coefficients to be accessed. This is assured as we have the correct number of independent variables and have eliminated only those rotations that do not affect the energy. However, a consequence of eliminating the

(18) (a) Raffenetti, R. C.; Ruedenberg, K. *Int. J. Quantum Chem. Symp.* **1970**, *3*, 625. (b) Hoffman, D. K.; Raffenetti, R. C.; Ruedenberg, K. *J. Math. Phys.* **1972**, *13*, 528.

(19) Koutecký, J.; Bonačić, V. *Chem. Phys. Lett.* **1971**, *10*, 401.

energetically redundant occupied–occupied and virtual–virtual rotations is that the optimum MO coefficients obtained will not be the canonical Roothaan coefficients. The latter may be obtained by diagonalizing the occupied–occupied and virtual–virtual blocks of the final Fock matrix in the MO basis

$$F_{pq} = \sum_{\mu\nu} C_{\mu p} F_{\mu\nu} C_{\nu q} \tag{17}$$

$$\sum_{kl} U^O_{ki} F_{kl} U^O_{lj} = e_i \delta_{ij} \tag{18}$$

$$\sum_{cd} U^N_{ca} F_{cd} U^N_{db} = e_a \delta_{ab} \tag{19}$$

The canonical coefficients, which of course yield exactly the same energy, are then given as a unitary transformation of the noncanonical MO's

$$C_{\mu i}' = \sum_{j} C_{\mu j} U^O_{ji} \tag{20}$$

$$C_{\mu a}' = \sum_{b} C_{\mu b} U^N_{ba} \tag{21}$$

The C' now satisfy the Roothaan equations (7) as well as $\partial E/\partial\theta = O$.

To perform a simultaneous optimization of geometry and wavefunction, we must still obtain an expression for the partial derivatives of the energy with respect to the geometric variables **r**. This will be done by finding the partial derivative of the energy with respect to a single Cartesian coordinate X of one of the nuclei, which generalizes to yield the gradient at every nucleus in cartesian coordinates, $\partial E/\partial X$. The Cartesian coordinates of the nuclei are a partly redundant set as translations, and rotations of all nuclei at once do not affect the molecular geometry. The gradient in Cartesian coordinates must therefore be transformed into nonredundant internal coordinates (such as bond lengths and angles) by using standard methods,[5] giving $\partial E/\partial\mathbf{r}$.

It should be stressed that the gradient $\partial E/\partial X$ derived below represents the true partial derivative of the energy with respect to X, subject to both the wave function variables θ and other geometric variables being held constant. By contrast, previous expressions[20] for the gradient in the force method represent the derivative subject to the condition (7) that the wave function is optimal at all geometries. Since the geometric variables and the wave function variables are independent, and the unitary transform **U** has no direct dependence on the geometric variables, then

$$\frac{\partial U(\theta)}{\partial X} = \sum_{ai} \frac{\partial U}{\partial\theta_{ai}} \frac{\partial\theta_{ai}}{\partial X} = O \tag{22}$$

Differentiating the energy expression (3) with respect to X, noting that the AO → OBF transformation **M** is a function of X, as are the nuclear-centered basis functions, leads to

$$\frac{\partial E}{\partial X} = 2 \sum_{\mu\nu wi} \frac{\partial M_{\mu w}}{\partial X} U_{wi} C_{\nu i} F_{\mu\nu} + \sum_{\mu\nu} P_{\mu\nu} \frac{\partial H_{\mu\nu}}{\partial X} +$$
$$2 \sum_{\mu\nu\lambda\sigma} P_{\mu\nu} P_{\lambda\sigma} \left(\frac{\partial\mu}{\partial X} \lambda \| \nu\sigma \right) + \frac{\partial V_{nuc}}{\partial X} \tag{23}$$

P is the spin-orbital density matrix, which is defined as

$$P_{\mu\nu} = \sum_{i} C_{\mu i} C_{\nu i} \tag{24}$$

Only the first term of the derivative expression (23) differs from the expression for $\partial E/\partial X$ with the condition of wave function optimality enforced.[20] In Appendix 2, the way **M** depends on molecular geometry to satisfy (8) is defined and its derivative with respect to X is obtained. By use of this result (equation A2-10 from Appendix 2), the derivative of the energy with respect to a nuclear coordinate X can be written in the standard form as a contraction of matrix element derivatives with known arrays

$$\frac{\partial E}{\partial X} = \sum_{\mu\nu} W_{\mu\nu} \frac{\partial S_{\mu\nu}}{\partial X} + \sum_{\mu\nu} P_{\mu\nu} \frac{\partial H_{\mu\nu}}{\partial X} +$$
$$2 \sum_{\mu\nu\lambda\sigma} P_{\mu\nu} P_{\lambda\sigma} \left(\frac{\partial\mu}{\partial X} \lambda \| \nu\sigma \right) + \frac{\partial V_{nuc}}{\partial X} \tag{25}$$

where **W** is an energy-weighted density matrix, defined as

$$W_{\mu\nu} = -2 \sum_{\lambda\sigma\eta\alpha\beta w i} \frac{V_{\eta w} U_{wi} C_{\sigma i} F_{\lambda\sigma} T_{\lambda\alpha} T_{\mu\alpha} T_{\nu\beta} T_{\eta\beta}}{s_\alpha^{1/2} s_\beta^{1/2} (s_\alpha^{1/2} + s_\beta^{1/2})} \tag{26}$$

T and **s** are the eigenvectors and eigenvalues of the overlap matrix **S**, while **V** is a unitary matrix obtained from **M** at the initial geometry, denoted by a superscript i (see Appendix 2 for further details).

$$V_{\mu x} = \sum_{\nu} (S^{+1/2})^i_{\mu\nu} M^i_{\nu x} \tag{27}$$

In the case of an optimized wave function satisfying the Roothaan equations (7) at the current geometry, **W** reduces to the energy-weighted density matrix given in previous work,[20] and therefore so too does the gradient expression.

3. Implementation

Our first-order algorithm for simultaneous optimization of geometry and wave function is conceptually straightforward:

(1) Input a starting geometry and specify an AO basis. Obtain an initial guess for the MO coefficients which will be used as **M**, and set $\theta = O$.

(2) Reorthogonalize **M(r)** if necessary. Calculate the energy $E(\mathbf{r},\theta)$ and its first partial derivatives $\partial E/\partial\mathbf{r}$ and $\partial E/\partial\theta$.

(3) Using an optimization algorithm which takes as input the function value (energy) and its first partial derivatives, generate another point $(\mathbf{r},\theta)$ at the predicted position of the minimum. If the current gradient and/or displacement length are still above some threshold, go back to (2).

If the wave function is to be optimized at a fixed geometry, then exactly the same algorithm is used, except that $\partial E/\partial\mathbf{r}$ is no longer computed.

This procedure has been incorporated into the Gaussian 86 MO program package,[21] for VAX computers, and the CSPI MAP-6430 array processor. A version to run on Cray supercomputers is in progress. Algorithms based on the Obara–Saika recurrence relations[22] are used for the evaluation of the two-electron integrals over Gaussian basis functions and their derivatives with respect to nuclear coordinates, which is the most time-consuming step. However, as this is not central to the joint optimization scheme, it will not be further discussed here.[23]

To obtain $\partial E/\partial\theta$ in step 2 of the above scheme, (14) for $\partial E/\partial U$ must first be evaluated. This requires generation of the Fock matrix, as in the Roothaan procedure, which requires $O(N^4)$ (meaning of order N^4) operations due to its dependence on the two-electron integrals. Subsequently, two matrix multiplies requiring $O(N^3)$ multiplies completes the formation of $\partial E/\partial U$.

$\partial E/\partial U$ is then combined with $\partial U/\partial\theta$ in (13) to yield $\partial E/\partial\theta$ as the $\partial U/\partial\theta_{ai}$ are formed, by the following "zipper" algorithm:

(a) Initialize: $U^{left} = 1$, $U^{right} = U$
(b) Loop over the OV θ_{ai}'s in the order defined in (11)

(c) $\qquad\qquad U^{right} \leftarrow G(\theta_{ai})^\dagger U^{right}$

This is an in-place matrix multiplication requiring only $O(N)$ floating point operations (Flops) due to the sparse structure of **G**.

(d) $\qquad \frac{\partial U}{\partial\theta_{ai}} \leftarrow U^{left} \frac{\partial G(\theta_{ai})}{\partial\theta_{ai}} U^{right}$

Only the first O columns of $\partial U/\partial\theta_{ai}$ are needed in (13), and exploiting the sparsity of $\partial G/\partial\theta_{ai}$ in (16) means that the above

(20) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. *Int. J. Quantum Chem. Symp.* **1979**, *13*, 225.

(21) Frisch et al. Carnegie-Mellon Quantum Chemistry Publishing Unit, Carnegie-Mellon University: Pittsburgh, PA 15213.
(22) Obara, S.; Saika, A. *J. Chem. Phys.* **1986**, *84*, 3963.
(23) Head-Gordon, M.; Pople, J. A., to be submitted for publication.

operation requires $O(ON)$ Flops.

(e) $\quad \dfrac{\partial E}{\partial \theta_{ai}} \leftarrow \sum_{wj} \dfrac{\partial E}{\partial U_{wj}} \dfrac{\partial U_{wj}}{\partial \theta_{ai}} \qquad O(ON)$ Flops

(f) $\qquad \mathbf{U}^{\text{left}} \leftarrow \mathbf{U}^{\text{left}}\, \mathbf{G}(\theta_{ai}) \qquad O(N)$ Flops

The total Flop count for this algorithm is $O(O^2VN)$, and is far less than the effort necessary to generate the two-electron integrals and Fock matrix. The evaluation of (13) alone requires $O(O^2VN)$ Flops when $\partial E/\partial U$ and $\partial U/\partial \theta_{ai}$ are dense matrices, meaning that it will not be possible to improve this order dependence (unless θ is reset to O on each iteration which is a change of variables, and would thus be incompatible with the use of quasi-Newton optimization methods). However, in the context of semiempirical MO methods, which are normally dominated by $O(N^3)$ operations since the Fock matrix is efficiently parameterized, the cost of forming $\partial E/\partial \theta$ would be a rate-limiting step, and this approach is unlikely to be competitive with existing alternatives.

Formation of partial derivatives of the energy with respect to Cartesian coordinates (25) is dominated by the calculation of $O(N^4)$ two-electron integral derivatives, which are formed concurrently with the integrals themselves. Altering the definition of the energy-weighted density matrix to (26) allows existing gradient programs to be used for this step. Although (26) appears cumbersome, it can be efficiently evaluated by a number of matrix multiplies and a scaling by the overlap eigenvalue expression. This procedure is dominated by the matrix multiplies, and accordingly requires only $O(N^3)$ Flops.

In the third step of our general procedure, we are currently using an optimization algorithm based on the popular Broyden–Fletcher–Goldfarb–Shanno (BFGS) method.[17,24] This is a quasi-Newton scheme for updating an approximate inverse Hessian (second derivative matrix) $\mathbf{H}^{-1}$ using the function and first derivative information from each point, so that it tends to the exact inverse as large numbers of points are obtained. Each new displacement Δx is found by predicting where the gradient will vanish, using Taylor series to first order, with the approximate $\mathbf{H}^{-1}$

$$\Delta x = -\mathbf{H}^{-1}\mathbf{g} \tag{28}$$

where $\mathbf{g}$ is the current gradient. To accommodate the extremely large numbers of independent variables (r,θ) in the HF problem, $\mathbf{H}^{-1}$ is not stored in the space of (r,θ), but rather in the small expanding subspace spanned by the gradients at each point sampled. Furthermore, once Δx has been obtained from (28) (in the subspace variables and then projected into the full space), no line search to optimize the displacement length is usually performed. However, if the energy rises from one point to the next, a quadratic curve is fitted through the two points and used to step backward to the predicted minimum along the old displacement direction.

The BFGS algorithm begins with a unit matrix for $\mathbf{H}^{-1}$. Accordingly it is advantageous to scale the independent variables (r,θ) in such a way that the diagonal second derivatives are close to unity. If, for example

$$\partial^2 E/\partial x_i^2 = k_i \tag{29}$$

then a scaled variable defined as

$$y_i = k_i^{1/2} x_i \tag{30}$$

can easily be shown by the chain rule to have a second derivative of 1. For second derivatives of the energy with respect to a given θ_{ai} in the limit of all θ's being zero, one can derive

$$\frac{\partial^2 E}{\partial \theta_{ai}^2} = 2(F_{aa} - F_{ii}) + 2(ai\|ia) \tag{31}$$

which is closely related to existing results from HF stability theory.[25] Estimates of the diagonal second derivatives are obtained

TABLE I: Single-Point 3-21G Calculations at the STO-3G Equilibrium Geometry

	water	methanol
unitary transform scheme iterations	9	10
Roothaan-based methods iterations	9	10
final energy/hartrees	−75.583 69	−114.393 42

TABLE II: 3-21G Geometry and Wave Function Optimizations

	water	methanol
simultaneous optimization iterations	16	26
force method		
geometry iterations	5	7
wave function iterations	48	76
total iterations	53	83
final energy/hartrees	−75.585 96	−114.395 66

in our program by transforming the initial Fock matrix to the basis of the initial guess MO's and diagonalizing it. The eigenvectors are used to define the OBF, and the corresponding eigenvalues are then used to approximate (31) as

$$\partial^2 E/\partial \theta_{ai}^2 \approx 2(e_a - e_i) \tag{32}$$

The use of scaled coordinates via (32) improves the initial convergence of the wavefunction considerably in conjunction with the BFGS method. Empirical estimates of diagonal force constants for geometric variables are already available within Gaussian 86, and these are utilized to scale the geometric variables.

4. Preliminary Applications

As an initial application of the new SCF procedures introduced here, calculations on water and methanol were performed with the split valence 3-21G basis.[26] Calculations on these molecules were also performed with the standard algorithms in the Gaussian 86 program, for comparison. The point that will be emphasized is the number of iterations necessary to converge the procedure, rather than the computer time needed. Computer processing time is highly dependent on the implementation of the rate-limiting computational steps, such as two-electron integral and derivative evaluation, rather than on the general structure of the algorithm. Even comparison of the iteration counts is by no means definitive, since, for example, it depends crucially on the choice of optimization algorithm used in the third step of the procedure outlined in section 3 (which is currently the BFGS method). In existing methods, it depends on the particular extrapolation scheme for converging the Roothaan procedure (which is Pulay extrapolation[9] in Gaussian 86) and the geometry optimization algorithm employed (the scheme due to Schlegel[27] in this case). Nevertheless, the results will still allow some preliminary assessment of the promise of these new procedures to be made.

The first set of results are summarized in Table I and are 3-21G wave function calculations at a fixed geometry, namely, the equilibrium geometry found with the minimal STO-3G basis.[28] This allows a comparison of the Roothaan method augmented by Pulay extrapolation, with the unitary transform scheme proposed here, used in conjunction with the BFGS method. All calculations begin with an initial guess which is a converged semiempirical wave function[29] that has been projected into the space of the 3-21G basis. The threshold for convergence is 10^{-9} hartree on the energy. It is evident from Table I that both methods converge quickly in a comparable number of iterations. This suggests that the unitary

(24) (a) Broyden, C. G. *J. Inst. Math. Its Appl.* **1970**, *6*, 76. (b) Fletcher, R. *Comput. J.* **1970**, *13*, 317. (c) Goldfarb, D. *Math. Comput.* **1970**, *24*, 23. (d) Shanno, D. F. *Math. Comput.* **1970**, *24*, 647.

(25) Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1977**, *66*, 3045.
(26) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.
(27) Schlegel, H. B. *J. Comput. Chem.* **1982**, *3*, 214.
(28) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657.
(29) Pople, J. A.; Beveridge, D. L.; Dobosh, P. A. *J. Chem. Phys.* **1967**, *47*, 2026.

transform SCF may be competitive with Roothaan-based methods for single point calculations; we are investigating this by studying a wider range of molecules.

The second set of results are optimizations of both geometry and wave function, using the 3-21G basis, beginning from the STO-3G geometry, and a semiempirical initial guess. Table II summarizes the results obtained by using both our simultaneous optimization procedure and the traditional force method, with the energy converged to 10^{-9} hartree, although the wave functions in the force method are converged slightly more accurately to ensure reliable gradients. For the force method, the number of geometry changes, the number of wave function changes (representing the full optimization of the wave function at each trial geometry) and the total number of iterations are given. Clearly the simultaneous optimization procedure converges in significantly fewer total iterations; in fact, it requires little more than one-third as many cycles as the force method. However, 3-4 times as many geometry changes are required.

Compared to the disk-based algorithms, in which the two-electron integrals are stored, the simultaneous optimization procedure looks inefficient. This is because each time the geometry changes, the two-electron integrals and derivatives must be recalculated, in the most time-consuming step of the procedure. By contrast, the many wave function iterations in the force method are relatively inexpensive, as the stored two-electron integrals are utilized. On the basis of the number of geometry changes, it thus appears that the simultaneous optimization procedure is currently 3-4 times slower than the force method.

This balance will change significantly in the context of the direct SCF method, where the integrals are recalculated on each wave function iteration, making such iterations much more time-consuming. Here the reduction in total iteration count will be much more beneficial, and it is likely that simultaneous optimization will be quite competitive. For example, in the methanol calculations, the two-electron integral evaluation time is about half the integral derivative evaluation time, meaning that the disjoint optimization is equivalent to roughly $7 \times {}^2/_3 + 76 \times {}^1/_3 \approx 29$ simultaneous optimization cycles. With further development of the computational techniques and optimization algorithms,[23] it therefore seems likely that the simultaneous optimization technique will be a useful extension of the direct SCF method for large molecule geometry optimizations.

Appendix 1. Parameterization of the Unitary Transformation

In this appendix, an exponential form for the unitary transformation is considered and compared with the Givens transformation method adopted in section 2. **U** is expressed as

$$\mathbf{U} = e^{\mathbf{X}} \qquad (A1\text{-}1)$$

where **X** must be anti-Hermitian

$$\mathbf{X} = -\mathbf{X}^\dagger \qquad (A1\text{-}2)$$

to ensure unitarity. The OV elements of **X** coupling occupied and unoccupied OBF's are taken as variables and indeed are the only nonzero elements of **X**.

In the limit as $\mathbf{X} \to \mathbf{O}$, analytic expressions for the derivatives of the energy with respect to X_{ai} have been derived and used to minimize the energy both in quadratically convergent second-order schemes[10,11] and first-order schemes.[12] The first derivatives obtained in this way are identical with our expressions for $\partial E / \partial \theta$ is the limit $\theta \to \mathbf{O}$. The requirement $\mathbf{X} \to \mathbf{O}$ means that the matrix **M** in (9) must be redefined on each cycle, which is equivalent to a *change of variables*. In connection with first-order methods, it is therefore not possible to accumulate approximate second derivative information via quasi-Newton optimization algorithms. Accordingly some kind of exact or approximate second derivative calculation is desirable to ensure good convergence.

Alternatively, to avoid costly second derivative calculations and to allow the use of quasi-Newton optimizers, one must evaluate the derivative of the energy with respect to X_{ai} when **X** is nonzero. Equations 13 and 14 of section 2 are valid for any definition of **U**, which just leaves the problem of obtaining $\partial U / \partial X_{ai}$. The exponential of a matrix is defined as a Maclaurin series, so that **U** is explicitly given by

$$\mathbf{U} = e^{\mathbf{X}} = \mathbf{1} + \mathbf{X} + \frac{1}{2!}\mathbf{XX} + \frac{1}{3!}\mathbf{XXX} + \dots \qquad (A1\text{-}3)$$

Therefore its derivative with respect to X_{ai} is

$$\frac{\partial \mathbf{U}}{\partial X_{ai}} = \mathbf{Y} + \frac{1}{2!}(\mathbf{YX} + \mathbf{XY}) + \frac{1}{3!}(\mathbf{YXX} + \mathbf{XYX} + \mathbf{XXY}) + \dots$$

$$(A1\text{-}4)$$

where $\mathbf{Y}(X_{ai})$ is defined as

$$Y_{pq} = \frac{\partial X_{pq}}{\partial X_{ai}} = \delta_{ap}\delta_{iq} - \delta_{ip}\delta_{aq} \qquad (A1\text{-}5)$$

Equations A1-4 and A1-5 are the analogues of (15) and (16) for this different definition of **U**.

The amount of work necessary to evaluate (13), (14), and (A1-4) is proportional to O^2VN, as it was in section 3 for the Givens transformations. This follows first from the fact that (13) requires $O(O^2VN)$ Flops (like step e of the zipper algorithm), since $\partial U / \partial X_{ai}$ in (A1-4) is a dense matrix. Second the individual terms of (A1-4) require $O(ON)$ Flops in the worst case (terms such as **XYX** which represent a triple product between a dense, a sparse, and a dense matrix), meaning an overall effort of $O(O^2VN)$ since (A1-4) and (A1-5) must be evaluated for each X_{ai}. This is akin to step d of the zipper algorithm.

We prefer the Givens transformation based representation of the unitary transform **U** over this exponential parameterization, due to the benefits of being able to represent **U** and its derivatives exactly in terms of trigonometric functions, rather than the matrix power series which arise in (A1-3) and (A1-4). The power series must be truncated at some finite order, which will increase with the magnitude of the elements of **X**, making the gradient expression (A1-4) increasingly cumbersome to evaluate. Products such as **XX**, **XXX**, etc. must be precomputed before the individual $\partial U / \partial X_{ai}$ are evaluated (to allow each $\partial U / \partial X_{ai}$ to be evaluated with only $O(ON)$ Flops) leading to growing storage requirements, as well as more arithmetic operations as higher terms are kept in (A1-4). Therefore we expect the exponential form of **U** to generally require a larger multiple of O^2VN Flops and to be more susceptible to round-off errors due to series truncation than the Givens-based method we have used.

Appendix 2. Derivative of the AO → OBF Transform with Respect to Geometry

In this appendix, an expression is obtained for the derivative of the orthonormalizing transformation **M** with respect to a Cartesian coordinate X. We make no assumptions as to how **M** is initially obtained, but we must specify how **M** changes with geometry so as to always satisfy the orthonormalization condition (8). Various methods are available to reorthogonalize the basis when the geometry and hence the AO overlap matrix change. We use symmetric orthogonalization[30] since the reorthogonalized vectors are closest to their nonorthogonal counterparts in a least-squares sense. It has already been utilized in obtaining energy derivatives when multiconfigurational SCF wave functions are employed.[31,32]

One limitation of symmetric orthogonalization, to be discussed in more detail elsewhere[23] along with an assessment of alternative orthogonalization procedures, is that the resulting OBF are not invariant to a rigid rotation of the molecular framework. Consequently, when the molecular geometry is changed by altering

(30) Lowdin, P. O. *Adv. Phys.* **1956**, *5*, 1.
(31) Jorgensen, P.; Simons, J. *J. Chem. Phys.* **1983**, *79*, 334.
(32) Helgaker, T.; Almlöf, J. *Int. J. Quantum Chem.* **1984**, *26*, 275.

the internal coordinates, it is important to orient the molecule in Cartesian space in a way which ensures that such changes are not accompanied by significant rigid rotation. For example, one could choose the principal axes of positive charge (which are analogous to the principal inertial axes, but with nuclear charges replacing masses in the definition of the inertia tensor) as the Cartesian axes. Alternatively, the so-called Z matrix orientation used in the Gaussian 86 program[21] has proven satisfactory in this regard and is employed for the examples reported in section 4.

The symmetric orthogonalization procedure we use is as follows. If the overlap matrix is **S**, with eigenvectors **T** and eigenvalues **s**

$$\sum_{\mu\nu} T_{\mu\alpha} S_{\mu\nu} T_{\nu\beta} = s_\alpha \delta_{\alpha\beta} \tag{A2-1}$$

then the matrix

$$(S^{-1/2})_{\mu\nu} = \sum_\alpha T_{\mu\alpha} s_\alpha^{-1/2} T_{\nu\alpha} \tag{A2-2}$$

can easily be shown to orthonormalize the basis via (8). The matrix **M** which we use to orthonormalize the basis can then be expressed as a unitary transform **V** of $S^{-1/2}$, since unitary transformations preserve orthonormality:

$$M_{\mu x} = \sum_\nu (S^{-1/2})_{\mu\nu} V_{\nu x} \tag{A2-3}$$

We determine **V** by inverting equation A2-3 at the initial geometry (indicated by a superscript i), using the given matrix of initial guess MO coefficients $\mathbf{M}^i$ and the initial overlap matrix

$$V_{\mu x} = \sum_\nu (S^{+1/2})^i_{\mu\nu} M^i_{\nu x} \tag{A2-4}$$

The unitary matrix **V**, once formed by (A2-4), is constant, and the way **M** changes with geometry is then related by (A2-2) to the way $S^{-1/2}$ changes.

To obtain the derivative of **M** with respect to X, an infinitesimal change Δ is applied to the geometry, and Taylor series expansions are used to find the first order response in **M**, $\partial M/\partial X$. For brevity this will be written as $\mathbf{M}^X$. The procedure is to expand everything which depends on Δ in equations A2-1 to A2-3, in powers of Δ, and then satisfy these perturbed equations to first order in Δ (they will be trivially satisfied at zero order). Thus

$$S_{\mu\nu} = S^0_{\mu\nu} + \Delta S^X_{\mu\nu} + \dots$$

$$T_{\mu\alpha} = \sum_\beta T^0_{\mu\beta}(\delta_{\beta\alpha} + \Delta J_{\beta\alpha} + \dots)$$

$$s_\alpha = s^0_\alpha + \Delta s^X_\alpha + \dots \tag{A2-5}$$

where the superscript O refers to unperturbed quantities. Factoring the expansion of **T** as above simplifies the subsequent algebra considerably. Note also that since only the derivatives of the AO overlap matrix $S^X_{\mu\nu}$ can be calculated directly, expressions for the other derivatives will therefore be cast in terms of the $S^X_{\mu\nu}$.

Substituting the expansion for **T** from (A2-5), into the unitarity condition (10) which it must satisfy, and collecting terms of first order gives

$$J_{\beta\alpha} = -J_{\alpha\beta} \tag{A2-6}$$

The X-dependent quantities in (A2-1) are now expanded in powers of Δ. When the terms first order in Δ are collected and simplified by using (A2-6), the results from the diagonal and off-diagonal equations are

$$s^X_\alpha = S^X_{\alpha\alpha}$$

$$J_{\alpha\beta} = -S^X_{\alpha\beta}/(s^0_\alpha - s^0_\beta) \tag{A2-7}$$

where $S^X_{\alpha\beta}$ is the derivative of the overlap integral matrix in the basis of the eigenvectors of the unperturbed overlap matrix

$$S^X_{\alpha\beta} = \sum_{\mu\nu} T^0_{\mu\alpha} S^X_{\mu\nu} T^0_{\nu\beta} \tag{A2-8}$$

Using the expressions (A2-7) for the perturbed eigenvalues and eigenvectors, it is now possible to obtain an expression for $(S^{-1/2})^X$ by expanding (A2-2) in powers of Δ. Collecting and simplifying the first-order terms leads to

$$(S^{-1/2})^X_{\mu\nu} = \sum_{\alpha\beta} \frac{-T^0_{\mu\alpha} S^X_{\alpha\beta} T^0_{\nu\beta}}{s^{01/2}_\alpha s^{01/2}_\beta (s^{01/2}_\alpha + s^{01/2}_\beta)} \tag{A2-9}$$

which, since **V** is constant in (A2-3) is directly related to $\mathbf{M}^x$ by

$$M^X_{\mu x} = \sum_\nu (S^{-1/2})^X_{\mu\nu} V_{\nu x} \tag{A2-10}$$