Theor Chem Acc (1997) 97:96–109

Theoretical
Chemistry Accounts

Extremely localized molecular orbitals (ELMO): a non-orthogonal Hartree-Fock method

Marc Couty[1], **Craig A. Bayse**[2], **Michael B. Hall**[2]

[1] Laboratoire de Chimie Théorique, Université de Marne-la-Vallée, 2 rue de la Butte Verte, F-93166 Noisy-le-Grand Cedex, France
[2] Department of Chemistry, Texas A&M University, College Station, TX 77843-3255, USA

Received: 30 December 1996 / Accepted: 5 June 1997

Abstract. A new optimization method for extremely localized molecular orbitals (ELMO) is derived in a non-orthogonal formalism. The method is based on a quasi Newton-Raphson algorithm in which an approximate diagonal-blocked Hessian matrix is calculated through the Fock matrix. The Hessian matrix inverse is updated at each iteration by a variable metric updating procedure to account for the intrinsically small coupling between the orbitals. The updated orbitals are obtained with approximately n^2 operations. No n^3 processes such as matrix diagonalization, matrix multiplication or orbital orthogonalization are employed. The use of localized orbitals allows for the creation of high-quality initial "guess" orbitals from optimized molecular orbitals of small systems and thus reduces the number of iterations to converge. The delocalization effects are included by a Jacobi correction (JC) which allows the accurate calculation of the total energy with a limited number of operations. This extension, referred to as ELMO(JC), is a variational method that reproduces the Hartree-Fock (HF) energy with an error of less than 2 kcal/mol for a reduced total cost compared to standard HF methods. The small number of variables, even for a very large system, and the limited number of operations potentially makes ELMO a method of choice to study large systems.

Key words: Extremely localized molecular orbitals – Non-orthogonal Hartree-Fock – Linear scaling

1 Introduction

Performing Hartree-Fock (HF) calculations for large systems is possible today using the direct-SCF techniques [1] developed over the last 10 years. Several attempts to reduce the computation time have been designed by addressing both the Fock matrix computation and the orbital updating techniques. The integral pre-screening method [2] of Almlöf is one of the most efficient ways to discard small two-electron integrals before calculating them, thus reducing the n^4 dependency over the Gaussian basis set dimension [3, 4]. Other techniques to reduce the Fock matrix calculation time have been based on a classical approach to evaluate the Coulomb operator in terms of multipolar expansion [5] and derivation of new threshold criteria to further reduce the computation of negligible two-electron integrals for the exchange operator [6]. With the use of the fast multipole methods [7] and complicated tree algorithms [8], the Fock matrix computation time has been reduced to a less than n^2 scaling and allows one to compute Fock matrices for large systems using several thousand basis functions. Other authors have tried to reduce the iteration time in the SCF algorithm by suppressing the n^3 diagonalization process of the Fock matrix [9]. Though the idea is without question appealing, the ultimate success in its implementation is yet to be achieved, since a large set of linear equations must be solved. Others have attempted to develop a one-electron effective Hamiltonian [10] with effective potentials by taking advantage of a localized picture of bond-type orbitals in alkanes. Various methods have been implemented in semiempirical calculations such as a divide-and-conquer technique [11] and a localized molecular orbital pseudo-diagonalization method [12]. This work represents the first stages of the development of a non-orthogonal extremely localized HF method which is intended to provide fast, accurate results for large systems.

A trivial way to reduce the total time to perform an HF calculation is to reduce the number of iterations to reach convergence. A robust and efficient method with a fast convergence rate is a strong criterion, and its efficiency is enhanced if the starting point is close to the solution (i.e., the initial "guess" vectors have been well-designed for a particular problem). Here, we present a new method to perform accurate approximate HF calculations with a perspective of efficiency (efficiency in convergence and generation of a quality starting point) through the use of localized non-orthogonal orbitals.

Correspondence to: M. Couty

Since the early days of HF theory [13], orthogonal orbitals have been used to minimize the energy. The a priori constraint that orthogonality represents has been circumvented by the introduction of the Fock operator. The HF orbitals are eigenvectors of the Fock operator and thus form an orthogonal set of orbitals. By using that property, the standard convergence method obtains the updated orbitals as eigenvectors of a Fock operator and thus maintains orthogonality. However, the HF energy is not a function of the orbitals individually but of the space they engender. Indeed, transforming an orthogonal to a non-orthogonal set of orbitals within the HF space does not change the energy. The drawback of this transformation is the increased complexity of the formula to calculate the energy. The development of post-HF theories has also used orthogonal orbitals as the simplicity of the formulas allows an efficient calculation of the Hamiltonian matrix elements. In MCSCF calculations, the orthogonality conditions have been introduced implicitly (as opposed to the introduction of Lagrange multipliers) through unitary transformations applied to an orthogonal basis set of orbitals [14]. Only a few groups have faced the complexity of non-orthogonal orbitals in general multiconfiguration calculations [15]. Non-orthogonality reduces the need for large CI expansions, and gives a more chemical and interpretative picture in ab initio calculations. The chemical sense of orbitals has always been present in theoretical chemistry either in the group theoretical or the localized representation.

In the first part of this work, we discuss alternate formulations of the SCF algorithm. The second part is devoted to the derivation of a localized non-orthogonal orbital HF algorithm and its potential for use in large systems. Some results showing the accuracy, but not the scalability, of the method are presented in the third part. All derivations are performed for closed-shell HF wave functions but can be adapted to ROHF or UHF wave functions with minor changes.

2 Theoretical development of ELMO

For N electron pairs, the HF wave function Ψ can be written as a Slater determinant built with the HF orbitals:

$$\Psi = |\varphi_1\overline{\varphi}_1 \ldots \varphi_N\overline{\varphi}_N| \ . \tag{1}$$

If the orbitals are taken to be orthogonal, the total HF electronic energy has the following form:

$$E = \sum_{i=1}^{N} 2h_i + \sum_{i,j=1}^{N} 2J_{ij} - K_{ij} \ , \tag{2}$$

where h_k, J_{kl} and K_{kl} are defined as:

$$h_k = \int_1 \varphi_k(1)\left(T - \sum_\alpha \frac{Z_\alpha}{r_{\alpha 1}}\right)\varphi_k(1)\, d\tau_1$$

$$J_{kl} = \iint_{1\ 2} \varphi_k(1)\varphi_k(1)\frac{1}{r_{12}}\varphi_1(2)\varphi_1(2)\, d\tau_1\, d\tau_2 \tag{3}$$

$$K_{kl} = \iint_{1\ 2} \varphi_k(1)\varphi_1(1)\frac{1}{r_{12}}\varphi_k(2)\varphi_1(2)\, d\tau_1\, d\tau_2 \ .$$

When a set of Lagrange multipliers is introduced to account for the orthogonality conditions, a new function, E', of the orbitals is defined.

$$E' = E - \sum_{i,j=1}^{N} \varepsilon_{ij}(\langle\varphi_i|\varphi_j\rangle - \delta_i^j) \ . \tag{4}$$

The HF orbitals which minimize E' satisfy Eq. (5)

$$\forall k = 1, N$$

$$\frac{\partial E'}{\partial \varphi_{jk}} = 0 \tag{5}$$

$$F\varphi_{jk} = \sum_{l=1}^{N} \varepsilon_{kl}\varphi_{jl} \ ,$$

where F is the Fock operator.

Thus, the HF orbital space is invariant for the Fock operator, but this condition does not guarantee that the solution will give the minimum energy. Because F is a function of the HF orbitals, the problem is solved self-consistently. Let us take the particular case of a finite basis set and the matrix equation form of the Roothaan-Hall equations [13]. Let $\mathbf{C}$ be the rectangular matrix of the optimized orbital coefficients in an orthogonal basis set, then Eq. (5) can be written as:

$$\mathbf{FC} = \mathbf{C}\varepsilon \ , \tag{6}$$

where the Fock matrix $\mathbf{F}$ is a function of $\mathbf{C}$ say $\mathbf{F}(\mathbf{C})$.

The usual way to solve Eq. (6) is to use an implicit iterative scheme (fluid dynamics vocabulary) as follows:

$$\mathbf{F}(\mathbf{C}^n)\mathbf{C}^{n+1} = \mathbf{C}^{n+1}\varepsilon^n \ , \tag{7}$$

$\mathbf{C}^n$ representing the matrix $\mathbf{C}$ at iteration n. In simple cases, this method is reliable and (in general) stable. In cases where the method is unstable, level-shifting [16a] and direct inversion of the iterative subspace (DIIS) [16b] can be employed to converge the calculation. However, applying this scheme also minimizes the quantity G:

$$G = \sum_{i=1}^{N}(h_i + \varepsilon_i) \ , \tag{8}$$

in which ε_i are the N lowest eigenvalues of $\mathbf{F}$. The quantity G is equal to the total energy if one uses the HF orbitals. For some particular systems (especially those containing transition metals), it happens that the implicit scheme produces a set of orbitals (sometimes extremely close to the HF orbitals) for which G is lower than $\mathbf{E}_{\mathrm{HF}}$. At that point, the implicit scheme diverges since it minimizes G and can never go up to reach the HF energy. There are numerous examples of systems in which a "broken symmetry" solution apparently has a lower energy [17]. The problem disappears with the implementation of level-shifters but it often requires 15 iterations to converge using a set of guess orbitals that would give an energy less than 1 kcal/mol above the HF energy. Thus, the implicit scheme (with level-shifters or

DIIS implementation) does not take full advantage of being initialized with a well-designed set of molecular orbitals (MOs), because several iterations are needed for DIIS to develop an approximate Hessian of sufficient accuracy to converge to a proper solution.

Another manner of solving Eq. (2) is an explicit iterative scheme, similar to gradient methods:

$$\mathbf{F}(\mathbf{C}^n)\mathbf{C}^n = \mathbf{C}^{n+1}\varepsilon^n \ . \tag{9}$$

Because the Fock matrix has large positive eigenvalues, this method is extremely unstable, and diverges rapidly. However, using a low-order Krylov-space expansion built upon $\mathbf{C}^n$ [18], it is possible to determine an accurate approximation of the solution of the implicit scheme described above.

The introduction of Lagrange multipliers to maintain the orthogonality constraints tremendously complicates gradient techniques compared to performing an implicit iterative scheme and requires several n^2N (where n is the number of basis functions) processes if one only solves for the occupied orbitals.

Generally efficient gradient methods have only been used while implicitly integrating the orthogonality conditions. For example, defining the variables as the matrix elements of a unitary transformation applied to an orthogonal set of orbitals leads to a method that requires several n^3 processes such as matrix products, orbital orthogonalization and resolution of large matrix equations.

An alternative to this paradoxical situation is to remove the orthogonality constraints and use efficient gradient techniques from numerical analysis. The price is, a priori, more complex formulae to gain efficient convergence. Using Lowdin's derivation [19], the total electronic energy using non-orthogonal orbitals (φ_i), $i = 1, N$ in the HF wave function Ψ is:

$$\Psi = |\varphi_1\overline{\varphi}_1 \ldots \varphi_N\overline{\varphi}_N| \tag{10}$$

$$E = \sum_{i,j=1}^{N} 2h_{ij}S^{-1}_{ij} + \sum_{i,j,k,l=1}^{N} (ij|kl)\left(2S^{-1}_{ij}S^{-1}_{kl} - S^{-1}_{ik}S^{-1}_{jl}\right) \ , \tag{11}$$

where S^{-1} is the inverse of the overlap matrix between the MOs. S_{ij}, h_{ij}, and $(ij|kl)$ are defined as follows:

$$h_{ij} = \int_1 \varphi_i(1)\left(T - \sum_\alpha \frac{Z_\alpha}{r_{\alpha 1}}\right)\varphi_j(1)\, d\tau_1$$

$$(ij|kl) = \int_1 \int_2 \varphi_i(1)\varphi_j(1)\frac{1}{r_{12}}\varphi_k(2)\varphi_1(2)\, d\tau_1\, d\tau_2 \tag{12}$$

$$S_{ij} = \langle\varphi_i|\varphi_j\rangle \ . \tag{13}$$

These formulae can be used directly to minimize the energy with respect to the natural variables of the problem: the MO coefficients in the basis sets.

The number of variables is then equal to nN. With the orthogonality constraints, there are $nN - N(N+1)/2$ independent variables, and $N(n-N)$ independent variables when using Levy's exponential transformation [14]. The number of variables becomes huge for a large system, particularly if the basis set is rich. For example, to study an alkane C_mH_{2m+2} with a triple-ζ + polarization basis set [20] (C:5s3p1d/H:3s1p), the total number of variables would be proportional to m^2 and would reach 1.2×10^6 for m equal to 100 with 3112 functions in the basis set.

Because the use of non-orthogonal orbitals is not a restriction, one can achieve the goal of obtaining chemically intuitive orbitals by means of non-orthogonal localized orbitals. It can be argued that the useful information and the time saving from the use of spatial symmetry would then be lost. This argument can easily be overcome by the definition of large systems as systems without global symmetry (crystals and C_{60} are not considered as truly large systems from the point of view of this study due to their high symmetry). In this view, it would be equivalent to performing a HF calculation with orthogonal orbitals and localizing the converged orbitals with an ad hoc criterion to get an equivalent localization picture. Generally, localization techniques (such as Boys' method [21] or internal projection of ad hoc localized orbital [22]) calculate "localized" orbitals by transforming the HF orthogonal orbitals into (orthogonal or non-orthogonal) orbitals that satisfy a localization criterion under the constraint that the HF orbital space and the localized orbital space must be identical. Such a localization technique ensures that the energy is invariant and defines "localized" orbitals as orbitals that have a major component on a small number of basis functions. However, if an orbital is "localized" in this manner, it does not mean that the orbital is developed on a limited number of basis functions. Apart from the large coefficients, the orbital has small coefficients on all the basis functions and nothing ensures that these small coefficients contribute little to the total energy. For example in CH_4, localized non-orthogonal orbitals can be extracted from the HF orbitals by a least-square minimization of the coefficients of a σ_{CH} orbital on the CGTOs of the 3 H atoms not involved in the bond. After setting the small coefficients to zero, the recalculated total energy is 10.3 kcal/mol above the HF energy. With the constraint that the "localized" orbital space and the HF orbital space are identical, the "localized" orbitals are expanded over all the basis functions; thus they are still delocalized.

Every other definition of localization would change the HF orbital space and many result in a total energy higher than the HF energy for a given basis set. In this paper, a general *definition of localization* is to impose *a requirement that an orbital have non-zero elements over a limited number of basis functions*. Previous work implemented this type of localization by defining local basis sets for bonds or lone pairs [23] or a distance range for basis set cutoffs [24]. Similarly, in this work, extremely localized molecular orbitals (ELMOs) are fully optimized with the aforementioned localization constraint using linear expansion coefficients as variables. Since the localized and HF orbital spaces are a priori different, and the total energy (of a single configuration Slater determinant wave function) will be higher than the HF energy is a given basis set. Thus, this definition of localization reduces the number of intrinsic variables of the optimization process. For example, to study an alkane C_mH_{2m+2} with a triple-ζ + polarization basis set

[20] (C:5*s*3*p*1*d*/H:3*s*1*p*) and a two-atom bond localization pattern, the number of variables is proportional to m, and is equal to 9,312 for m equal to 100 and thus could be performed on a small computer. This definition is the central assumption in ELMO.

Using the above orbital localization definition, one way of optimizing the orbitals has been derived using a non-orthogonal configuration interaction with single excitations (CIS) procedure [25]. This method requires the evaluation of Hamiltonian matrix elements in a non-orthogonal set of Slater determinants. In the particular case of a single configuration with doubly occupied orbitals, the dimension of the CI matrix would be equal to $2N(n-N)+1$, a number which would restrict the method to small systems. This dimension can be reduced to $N(n-N)+1$ by applying a singlet spin-coupled scheme [26] but does not represent a fundamental change in the method. The two-electron integral transformation from atomic orbital (AO)-basis set to a MO-basis set in such a technique and the evaluation of a large number of non-orthogonal CI matrix elements are limiting factors to the use of this method in large systems.

2.1 ELMO orbital optimization method

Indices p, q, r, s and a, b will refer to the basis functions; indices i, j, k, l will refer to the occupied orbitals; n is the total number of basis functions; N is the number of occupied orbitals; and m is the total number of variables. The variables of the method are the orbital expansion coefficients stored in a matrix **C**:

$$\forall\, i = 1, N$$
$$\varphi_i = \sum_{p=1}^{n} C_{pi}\chi_p \tag{14}$$

where (χ_p), $p = 1, n$ are the basis functions.

The definition of the ELMO localization can be expressed by the fact that the matrix **C** is sparse. In such a case, the variables will be stored in a one-dimensional array **X**. The total energy can be written in terms of a density matrix ρ so that:

$$E = \sum_{p,q=1}^{n} 2h_{pq}\rho_{pq} + \sum_{p,q,r,s=1}^{n} (pq\,\|\,rs)\rho_{pq}\rho_{rs}$$
$$\rho_{pq} = \sum_{i,j=1}^{N} C_{pi}C_{qj}S_{ij}^{-1} \tag{15}$$
$$(pq\,\|\,rs) = 2(pq|rs) - (pr|qs)$$

with

$$h_{pq} = \int_1 \chi_p(1)\left(T - \sum_{\alpha}\frac{Z_\alpha}{r_{\alpha 1}}\right)\chi_q(1)\,d\tau_1$$
$$(pq|rs) = \int_1\int_2 \chi_p(1)\chi_q(1)\frac{1}{r_{12}}\chi_r(2)\chi_s(2)\,d\tau_1\,d\tau_2\ . \tag{16}$$

The energy calculation requires the evaluation of the matrix $\mathbf{S}^{-1}$, but this matrix has relatively small dimension (number of electron pairs) and can be calculated without difficulty even for a very large system. The matrix elements S_{ij} are calculated by:

$$S_{ij} = \sum_{p,q=1}^{n} C_{pi}C_{qj}\sigma_{pq} \tag{17}$$
$$\sigma_{pq} = \langle\chi_p|\chi_q\rangle\ .$$

The quantities σ_{pq}, h_{pq}, and $(pq|rs)$ depend only on the basis set and can be recalculated when required as in the direct-SCF technique. It can be noted that the evaluation of matrices **S** and ρ can be done with approximately Nn^2 operations (with an intermediate storage matrix) in the case that **C** is not a sparse matrix. However, the localization definition reduces the number of operations to n^*m with an additional storage, m being the total number of variables. Two matrices **A** and **B**, defined as follows, can be computed in a small number of operations if the sparse matrix **C** is used in its linear vector form **X**:

$$\forall\, k = 1, N$$
$$\forall\, q = 1, n$$
$$A_{kq} = \sum_{j=1}^{N} C_{qj}S_{kj}^{-1} \tag{18}$$
$$B_{kq} = \sum_{p=1}^{n} C_{pk}\sigma_{pq}\ .$$

The evaluation of **S** and ρ can also be obtained in a limited number of operations. Several important equations can be derived from the $\mathbf{SS}^{-1} = \mathbf{1}$ relation:

$$\sum_{q=1}^{n} B_{iq}A_{jq} = \delta_i^j$$
$$\sum_{p,q=1}^{n} \rho_{pq}\sigma_{pq} = N\ . \tag{19}$$

Interestingly, the matrix product $\sigma\rho$ can be computed in Nn^2 operations (instead of n^3) by using the relationship:

$$\sum_{k=1}^{N} B_{kp}A_{kq} = \sum_{r=1}^{n} \sigma_{rp}\rho_{rq}\ . \tag{20}$$

The localization pattern is conserved throughout the iterations by calculating only the gradient of the energy with respect to the variables and defining a step vector $\delta\mathbf{X}$ that has the same localization definition as the vector **X**.

A standard Newton-Raphson procedure is derived using a second-order Taylor series expansion at the current point **X**:

$$E(X+\delta X) = E(X) + \sum_{o=1}^{m} \mathrm{Grad}(X)_o\delta X_o + \frac{1}{2}\sum_{o,o'=1}^{m} \delta X_o \mathrm{Hess}(X)_{oo'}\delta X_{o'}\ . \tag{21}$$

The minimum of the second-order energy is obtained for $\delta\mathbf{X}$ that satisfies the matrix form of the Newton-Raphson equation:

$$\text{Hess}\ \delta X + \text{Grad} = 0 \ . \tag{22}$$

Though extremely efficient to converge in a limited number of iterations, this method requires the calculation of a large Hessian matrix (Hess) which can be very expensive even with the localization procedure. *Here, a quasi Newton-Raphson method is proposed, based on the fact that orbitals are strongly independent* (*and not connected by the orthogonality constraint*); *thus, the change in one orbital can be partially decoupled from the others*. A good approximation of the Hessian matrix can be developed by a diagonal orbital blocking (blocks of second derivatives corresponding to a given orbital). The Hessian coupling elements can be introduced numerically in the course of the iterations by finite differences of the accumulated gradients to achieve quadratic convergence. The only necessary elements to be derived are:

$$\begin{aligned} &\frac{dE}{dC_{ak}}(\mathbf{X}) \ , \\ &\frac{d^2E}{dC_{ak}\,dC_{bk}}(\mathbf{X}) \ , \end{aligned} \tag{23}$$

in which, C_{ak} and C_{bk} are variables of the problem.

The gradient and Hessian calculation require the first and second derivatives of the matrix elements ρ_{pq} and S^{-1}_{ij}.

First derivatives:

$$\frac{d\rho_{pq}}{dC_{ak}} = \sum_{i,j=1}^{N} C_{pi}C_{qj}\frac{dS^{-1}_{ij}}{dC_{ak}} + \delta^a_p A_{kq} + \delta^a_q A_{kp} \tag{24}$$

the relation $d\mathbf{S}^{-1} = -\mathbf{S}^{-1}\,d\mathbf{S}\,\mathbf{S}^{-1}$ allows one to calculate the first derivatives of $\mathbf{S}^{-1}$:

$$\frac{dS^{-1}_{ij}}{dC_{ak}} = -S^{-1}_{ik}\sum_{l=1}^{N} B_{la}S^{-1}_{lj} - S^{-1}_{jk}\sum_{l=1}^{N} B_{la}S^{-1}_{li} \tag{25}$$

and finally,

$$\frac{d\rho_{pq}}{dC_{ak}} = \delta^a_p A_{kq} + \delta^a_q A_{kp} - A_{kp}\sum_{l=1}^{N} B_{la}A_{lq} - A_{kq}\sum_{l=1}^{N} B_{la}A_{lp} \ . \tag{26}$$

The matrix $\mathbf{A}^* = {}^t\mathbf{BA}$ is introduced and can be computed in Nn^2 operations, stored in memory and its elements addressed when needed to compute the first derivative of the matrix elements of ρ:

$$\frac{d\rho_{pq}}{dC_{ak}} = \delta^a_p A_{kq} + \delta^a_q A_{kp} - A_{kp}A^*_{aq} - A_{kq}A^*_{ap} \tag{27}$$

Second derivatives:

$$\begin{aligned} \frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}} = &\left(\delta^a_p\delta^b_q + \delta^b_p\delta^a_q\right)S^{-1}_{kk} + \delta^a_p\sum_{j=1}^{N} C_{qj}\frac{dS^{-1}_{kj}}{dC_{bk}} \\ &+ \delta^a_q\sum_{j=1}^{N} C_{pj}\frac{dS^{-1}_{kj}}{dC_{bk}} + \sum_{i,j=1}^{N} C_{pi}C_{qj}\frac{d^2S^{-1}_{ij}}{dC_{ak}dC_{bk}} \\ &+ \delta^b_p\sum_{j=1}^{N} C_{qj}\frac{dS^{-1}_{kj}}{dC_{ak}} + \delta^b_q\sum_{j=1}^{N} C_{pj}\frac{dS^{-1}_{kj}}{dC_{ak}} \ . \end{aligned} \tag{28}$$

Another computational advantage can be achieved when calculating the matrix elements of the matrix product ${}^t\mathbf{BS}^{-1}$ corresponding to the couple (a,k) representing a variable. The full matrix product is thus avoided and a one-dimensional array can be stored. However, for simplicity of notation, the element $({}^t\mathrm{BS}^{-1})_{ak}$ will be denoted B^*_{ak} or in matrix notation

$$\mathbf{B}^* = {}^t\mathbf{BS}^{-1} \ . \tag{29}$$

The second derivative of $\mathbf{S}^{-1}$, $d^2\mathbf{S}^{-1}$ is calculated using:

$$\begin{aligned} \frac{d^2S^{-1}}{dXdY} = &S^{-1}\frac{dS}{dY}S^{-1}\frac{dS}{dX}S^{-1} + S^{-1}\frac{dS}{dX}S^{-1}\frac{dS}{dY}S^{-1} \\ &- S^{-1}\frac{d^2S}{dXdY}S^{-1} \ . \end{aligned} \tag{30}$$

Thus, the second derivative analytical expression becomes:

$$\begin{aligned} \frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}} = &\left(\delta^a_p\delta^b_q + \delta^b_p\delta^a_q\right)S^{-1}_{kk} \\ &- \delta^a_p\left(S^{-1}_{kk}A^*_{bq} + A_{kq}B^*_{bk}\right) - \delta^a_q\left(S^{-1}_{kk}A^*_{bp} + A_{kp}B^*_{bk}\right) \\ &- \delta^b_p\left(S^{-1}_{kk}A^*_{aq} + A_{kq}B^*_{ak}\right) - \delta^b_q\left(S^{-1}_{kk}A^*_{ap} + A_{kp}B^*_{ak}\right) \\ &+ A_{kp}\left(B^*_{ak}A^*_{bq} + B^*_{bk}A^*_{aq}\right) + A_{kq}\left(B^*_{bk}A^*_{ap} + B^*_{ak}A^*_{bp}\right) \\ &+ S^{-1}_{kk}\left(A^*_{bp}A^*_{aq} + A^*_{bq}A^*_{ap}\right) \\ &+ 2A_{kp}A_{kq}\left(\sum_{i,j=1}^{N} B_{ia}B_{jb}S^{-1}_{ij} - \sigma_{ab}\right) \ . \end{aligned} \tag{31}$$

The Fock matrix $\mathbf{F}$ can be introduced to accelerate the Hessian and gradient calculation:

$$F_{pq} = h_{pq} + \sum_{r,s}(pq\,\|\,rs)\rho_{rs} \ . \tag{32}$$

The gradient and Hessian formulas can be simplified to the following expressions:

$$\begin{aligned} &\frac{dE}{dC_{ak}}(\mathbf{X}) = \sum_{p,q} 2F_{pq}\frac{d\rho_{pq}}{dC_{ak}} \\ &\frac{d^2E}{dC_{ak}dC_{bk}}(\mathbf{X}) = \sum_{p,q} 2F_{pq}\frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}} \\ &\qquad + 2\sum_{p,q,r,s}(pq\,\|\,rs)\frac{d\rho_{pq}}{dC_{ak}}\frac{d\rho_{rs}}{dC_{bk}} \ . \end{aligned} \tag{33}$$

From the practical point of view, the Fock matrix is built up by accumulation of the contributions of the two-electron integrals. The gradient vector is then calculated by a n^2-process in which the first derivatives are calculated directly without any additional storage. The Hessian formula can be separated into two contributions. The first contribution, involving second derivatives of the density matrix, can be evaluated at low cost by following the gradient algorithm. The second contribution involving products of first derivatives is a priori the difficult part of the calculation since it cannot take advantage of any possible preliminary calculations (the first derivatives cannot be stored in memory for a large system).

2.2 The quasi Newton-Raphson algorithm

The algorithm proposed here is based on a calculation of an approximate blocked Hessian matrix of high quality using a Davidon-Fletcher-Powell (DFP) [27] or Broyden-Fletcher-Goldfarb-Shanno (BFGS) [27] updating technique in the course of the iterations. These two algorithms differ only numerically with the BFGS method considered the superior to DFP. The N blocked-Hessian matrix is approximated by the formula:

$$\frac{d^2E}{dC_{ak}dC_{bk}}(\mathbf{X}) = \sum_{p.q} 2F_{pq}\left(\frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}}\right)^* \tag{34}$$

in which:

$$\begin{aligned}\left(\frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}}\right)^* &= \left(\delta^a_p\delta^b_q + \delta^b_p\delta^a_q\right)S^{-1}_{kk} \\ &\quad - \left(\delta^a_pA^*_{bq} + \delta^a_qA^*_{bp} + \delta^b_pA^*_{aq} + \delta^b_qA^*_{ap}\right)S^{-1}_{kk} \\ &\quad + S^{-1}_{kk}\left(A^*_{bp}A^*_{aq} + A^*_{bq}A^*_{ap}\right) \\ &\quad + 2A_{kp}A_{kq}\left(\sum_{i,j=1}^{N} B_{ia}B_{jb}S^{-1}_{ij} - \sigma_{ab}\right) \quad . \end{aligned} \tag{35}$$

This particular choice has been made because of its simplicity since it requires only the Fock matrix and also because it has the same zero eigenvalues and eigenvectors as the true Hessian matrix (see Appendix). The fast rate of convergence demonstrated below justifies a posteriori the validity of this choice. An algorithm in N Newton-Raphson (N-R) steps can be easily implemented and used at a low cost until all the N-R steps can be completed successfully (a trust radius step restricted Fletcher method [28] is implemented in cases for which a negative eigenvalue of the Hessian matrix is encountered or if the N-R step goes beyond a trust radius). At that point, the blocked Hessian matrix is not recalculated but its inverse is updated using the DFP or BFGS algorithms thus taking into account numerically the density matrix first derivative contributions to the blocked Hessian as well as the inter-orbital Hessian matrix elements. The orbital decoupling hypothesis, efficient in the N-R first steps of the iterative process, may be problematic when close to convergence if several orbitals are localized in the same region of space. In that case, the DFP or BFGS updating technique must include numerically the coupling between the orbital variations and require saving only two (DFP) or three (BFGS) vectors every iteration. Using the DFP updating technique, the complete inverse Hessian matrix at iteration (i) is of the form:

$$H_i^{-1} = H^{-1}_{\text{blocked}} + \sum_{j=1}^{i-1}\alpha_j u_j \otimes u_j + \sum_{j=1}^{i-1}\beta_j v_j \otimes v_j \quad . \tag{36}$$

In the BFGS updating technique, another term of the form $\sum_{j=1}^{i-1}\gamma_j w_j \otimes w_j$ is added to the DFP inverse Hessian matrix.

The quasi Newton-Raphson step $\delta X = H_i^{-1}\text{Grad}$ is calculated in two separate steps. The diagonal block matrix-vector product can be done at a low cost in an N matrix-vector product of small size. Then, the updating vectors are retrieved sequentially and used to calculate a contribution to $\delta\mathbf{X}$ using the relation $(u \otimes u)v = (u \cdot v)u$ that requires only a scalar dot product of two vectors.

2.3 Convergence of the ELMO method

In order to test the ELMO algorithm convergence properties, the molecule H_2CO is taken as an example. A triple-ζ + polarization [20] (C, O:5s3p1d/H:3s1p) basis set is used at the HF-optimized geometry [29]. The initial localized MO guess is constructed with two-atom bond orbitals and corresponds to 110 variables (380 variables in a standard HF calculation) in the ELMO optimization procedure. The results using the DFP and BFGS updating techniques are presented in Tables 1 and 2. In order to compare the convergence of ELMO to the standard HF method convergence, the same initial guess is transformed to orthogonal orbitals and used as the initial guess in two HF calculations: one using the basic implicit scheme described previously and the other using the standard HF method with level shifters and DIIS. The results are presented in Tables 3 and 4. The standard implicit scheme convergence deficiencies are improved with the DIIS implementation. The convergence behavior of the DFP and BFGS are similar to one another and are comparable with the convergence of the DIIS procedure in standard HF methods. According to the systems studied to date, the BFGS updating technique can save a small number of iterations compared to the DFP algorithm. As expected, the convergence in the ELMO method is achieved significantly more precisely for the energy than the convergence for the gradient vector norm.

2.4 Influence of the localization constraints on the total energy and the density function

The localization constraint is evaluated by the energy difference between the HF and ELMO energies and summarized for four small systems in Table 5. This energy difference can be compared with the contribution of the orbital optimization to the total energy. One can

Table 1. Extremely localized molecular orbital (ELMO) calculation of formaldehyde using a quasi Newton-Raphson algorithm with the DFP updating technique

Iteration	Energy (a.u.)	Grad. norm	ΔE (a.u.)	Step norm
1	−113.873607	0.408580		0.107626
2	−113.896839	0.118891	−0.023232	0.051689
3	−113.898693	0.052097	−0.001854	0.002441
4	−113.898863	0.017386	−0.000169	0.000838
5	−113.898889	0.005365	−0.000026	0.000027
6	−113.898892	0.001845	−0.000003	0.000012
7	−113.898893	0.000861	−0.000001	0.000008
8	−113.898893	0.000358	−0.000000	0.000001
9	−113.898893	0.000090	−0.000000	

distinguish two different effects in an orbital optimization process. The first one is connected with the mixing of pure AOs. This effect is easily calculated by performing a HF-LCAO calculation in a minimal basis set made of the pure AOs of the free atoms, which are calculated with an extended basis set. The second effects, referred to as the orbital optimization effect [30], is connected with a change in the shape of the orbitals to satisfy various conditions such as the atomic cusp and the long range behavior of an MO. The difference between the HF and HF-LCAO is associated with the second effect. The relative difference between the optimization effect and the localization constraint is small in all cases (less than 7%) and will generally be a small effect. Particularly, the ELMO calculations performed in this work use the most localized orbitals one can think of i.e. *two-atom bonds* (this small energy constraint could even be decreased by extending the localization of a bond to include its closest neighbor atoms). In the H_2O molecule for example, the two non-bonding lone pairs were localized on the oxygen atom only. The localization constraints of the non-bonding orbital is responsible for a large amount of the energy difference with the HF energy.

Table 2. ELMO calculation of formaldehyde using a quasi Newton-Raphson algorithm with the BFGS updating technique

Iteration	Energy (a.u.)	Grad. norm	ΔE (a.u.)	Step norm
1	−113.873607	0.408580		0.107626
2	−113.896839	0.118891	−0.023232	0.057516
3	−113.898679	0.056213	−0.001809	0.003871
4	−113.898860	0.001148	−0.000181	0.001148
5	−113.898889	0.006057	−0.000027	0.000091
6	−113.898892	0.001790	−0.000003	0.000022
7	−113.898893	0.000643	−0.000001	0.000015
8	−113.898893	0.000324	−0.000000	0.000006
9	−113.898893	0.000099	−0.000000	

Table 3. Calculation of formaldehyde using a standard Hartree-Fock (HF) method with a basic implicit scheme

Iteration	Energy (a.u.)
1	−113.873607
2	−113.906958
3	−113.908887
5	−113.909227
10	−113.909310
15	−113.909320
20	−113.909321

Table 4. Calculation of formaldehyde using a standard HF method with level-shifters and direct inversion of the iterative subspace (DIIS) implementation

Iteration	Energy (a.u.)	Grad. norm	ΔE (a.u.)	
1	−113.873607	0.121372		Level-shifters
2	−113.898970	0.030026	−0.025361	”
3	−113.904458	0.025306	−0.005488	”
4	−113.907089	0.018429	−0.002631	”
5	−113.908227	0.013500	−0.001137	DIIS
6	−113.909287	0.002133	−0.001060	”
7	−113.909319	0.000292	−0.000032	”
8	−113.909320	0.000147	−0.000001	”
9	−113.909321	0.000039	−0.000000	”

In order to gain insight into the orbital optimization mechanism in the ELMO theory, the formaldehyde molecule H_2CO is taken as an example because the CH bond orbitals are defined in a unique way by the two-atom bond localization picture. The geometry was optimized at the HF level in a triple-ζ + polarization basis set [20]. From the MO expansion coefficients of the CH orbital, it has to be noted that the ratio between the coefficients relative to the p_x and p_z CGTOs in the same shell varies from shell to shell. One would have expected this ratio to be constant for all p shells and be a function only of the COH angle. A contour plot of the CH molecular orbital in the COH plane is given in Fig. 1 showing the distortion of the CH bond outside the bonding region. The two-atom ‘‘banana’’ bond in the cyclopropene has also been studied and a contour plot in the C_3 plane is given in Fig. 2.

A second comparison can be made on the density functions at the HF and ELMO levels. Because the total energies are similar, it is expected that the total densities may be comparable. To quantify the difference density function $(\rho_{\mathrm{HF}} - \rho_{\mathrm{ELMO}})$ and not its value on particular points, it is possible to calculate the variance between the densities as the integral [22]

$$|\rho_{\mathrm{HF}} - \rho_{\mathrm{ELMO}}| = \sqrt{\int\int\int_{\mathrm{space}} (\rho_{\mathrm{HF}} - \rho_{\mathrm{ELMO}})^2 \, d\tau}. \quad (37)$$

The results are given in Table 5 for small molecules. This distance is also compared with the distance between the HF and HF-LCAO density functions. The distance

Table 5. Comparison Between HF, HF-LCAO, ELMO and GMO2(PEMCSCF) total energies[a] for several small molecules

	CH_4	H_2O	NH_3	C_2H_6
Bond length (B)	CH = 2.04620	OH = 1.77746	NH = 1.88646	CC = 2.88638
E(HF) (a.u.)	−40.212074	−76.056015	−56.216625	−79.257272
E(ELMO) (a.u.)	−40.207443 (2.9)	−76.045341 (6.7)	−56.208059 (5.4)	−79.238197 (12.0)
E(HF-LCAO) (a.u.)	−40.048021 (102.9)	−75.904338 (95.2)	−56.032982 (115.2)	−79.985242 (170.7)
E(GMO2) (a.u.)	−40.272817 (−38.1)	−76.097563 (−26.1)	−56.269492 (−33.2)	−79.363848 (−66.9)
$\lvert\rho_{\mathrm{HF}}$-$\rho_{\mathrm{LCAO}}\rvert$	0.19046	0.18630	0.18679	0.23293
$\lvert\rho_{\mathrm{HF}}$-$\rho_{\mathrm{ELMO}}\rvert$	0.00401	0.01129	0.00819	0.00677
$\lvert\rho_{\mathrm{HF}}$-$\rho_{\mathrm{GMO2}}\rvert$	0.01575	0.01567	0.01468	0.02074

[a] Relative energies compared to HF are given in parentheses (kcal/mol)

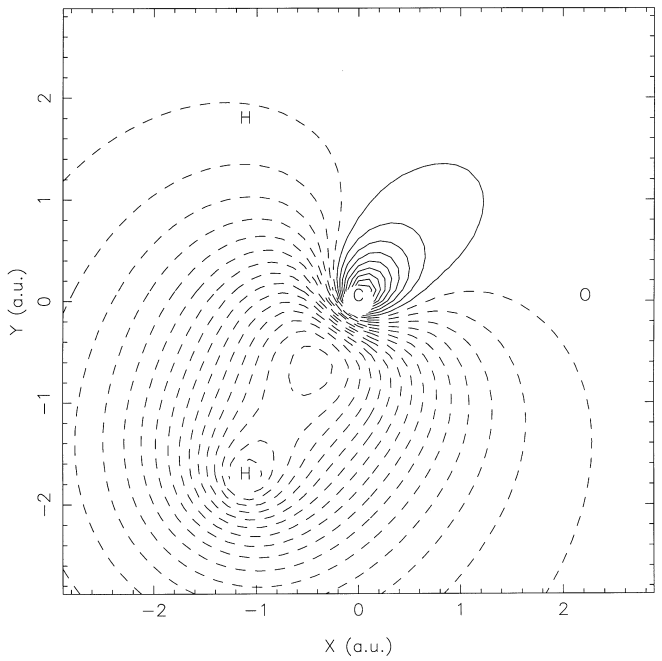

Fig. 1. Contour plot of the σ_{CH} two-atom bond of H_2CO in the plane of the molecule

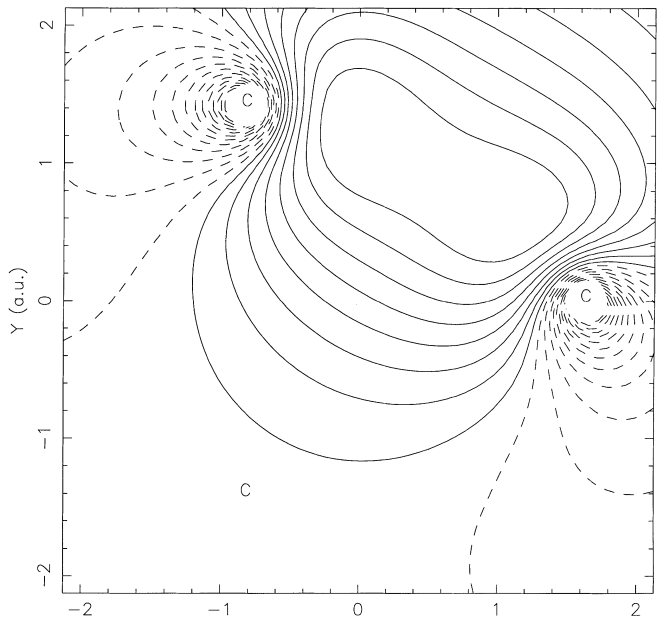

Fig. 2. Contour plot of the σ_{CC} two-atom bond of cyclopropane in the C_3 plane

between the ELMO and HF density functions is one order of magnitude smaller than the distance between the HF and HF-LCAO density functions. It has been noticed previously that the HF density and the density obtained with a calculation including a large amount of correlation energy are similar. This similarity has been used as a guiding force for density functional theories that use the HF density to calculate an accurate total energy. As a measure of the effect of electron correlation, the distance between the HF and GMO2(PEMCSCF) [31] density functions are also given in Table 5. GMO2(PEMCSCF) has been chosen for its simplicity. The optimized orbitals are the natural orbitals and do not necessitate any density matrix diagonalization. However, a reasonably large amount of "near-degenerate" correlation energy is obtained at this level of theory. The distance between the HF and ELMO density functions is then between 1 and 4 times shorter than the distance between the HF and GMO2(PEMCSCF). Thus, the correlation energy is large compared to the localization energy constraints.

The two-atom localized bond picture used in the ELMO calculations gives a very large fraction of the orbital optimization effects, reproducing the HF results (total energy and density function) with a relatively high accuracy. In this tight localization picture, the missing effects can be clearly expressed as the delocalization of the major component of a bond to other parts of the molecule.

2.5 Orbital delocalization

As a result of optimizing two-atom bond orbitals in ELMO, the total energy is higher than the HF energy. This energy difference, though small (2–20 kcal/mol), could be decreased by delocalizing the orbitals over the molecule and is zero in the limit of fully delocalized orbitals. However, the number of variables to be optimized increases drastically with the extension of the localization pattern. Since the delocalization energy (difference between the HF and the ELMO energies) is small, an approximate delocalization of the orbitals could be achieved by the diagonalization of the Fock matrix built using the converged ELMO orbitals. However, for large systems, this process cannot be performed at a low cost and an approximate treatment will be implemented.

From the Fock matrix built at the last iteration of the ELMO procedure, the canonical orbitals (that diagonalize the Fock matrix over the occupied orbitals and are linear combinations of the two-atom bond orbitals, thus containing no delocalization information) are calculated. The basis set is completed with an orthogonal set of MOs and the Fock matrix is easily transformed from the AO basis set to a MO basis set. An alternative using a low-order Krylov space [18] is currently under development. The Fock matrix has then the following pattern:

D	ε
ε	F'

in which D is diagonal. The Fock matrix elements ε_{ij} are intrinsically small and couple the occupied and virtual orbitals. The quantity κ is a measure of the Brillioun-violation. A small value reflects that the major bonding effects have been included at this stage, where κ is written as

$$\kappa = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{n-N}\varepsilon_{ij}^2} \quad . \tag{38}$$

It is then possible to obtain accurately the N functions corresponding to the N lowest eigenvalues by a Jacobi correction (JC) [32], which allows the delocalization of the orbitals. The JC (order Q) consists in a succession of Q Jacobi sweeps which reduce the value of the quantity κ. The method converges in the limit of infinite order to a result equivalent to diagonalization of the Fock matrix, but requires a smaller number of operations compared to the diagonalization of the entire matrix. In practice, a small number of passes are required for the method to approach convergence if the localization pattern of the ELMO procedure allows for an optimized treatment of the major bonding effects. In the following section this *sine qua non* condition will be examined in details. *Nota Bene*: It is not possible to obtain any useful reliable total energy stabilization from the orbital delocalization process only, since the eigenvalues of **F** do not properly account for the two-electron interaction energy. To remedy this intrinsic problem, the total energy is evaluated exactly by the actual computation of the Fock matrix with the delocalized orbitals. Using these delocalized orbitals, the density matrix is calculated and can be expressed as $\rho(\mathrm{ELMO}) + \delta\rho$ in which the elements of $\delta\rho$ are intrinsically small. The Fock matrix is evaluated efficiently, using a pre-screening on $\delta\rho$, as $\mathbf{F}(\mathrm{ELMO}) + \delta\mathbf{F}$, and the total energy is calculated at low cost by the usual formula in a n^2 process:

$$E = Tr[(h + F)\rho] \ . \tag{39}$$

Though the delocalized ELMO orbitals are obtained by a Jacobi process, the method is variational and one can state that the total energy is higher than the HF energy. Application of ELMO(JC) to various small systems is shown in Table 6. The difference between the HF and ELMO(JC) total energies ranges from 0.0 to 1.0 kcal/mol and is, in all cases, extremely small compared to the atomization energy.

Table 6. Comparison between ELMO, ELMO(JC) and HF total energies[a] for various small systems

	ELMO (a.u.)	ELMO(JC) (a.u.)	HF (a.u.)
CH_4	−40.20744 (2.9)	−40.21197 (0.06)	−40.21207
NH_3	−56.20806 (5.4)	−56.21643 (0.12)	−56.21662
H_2O	−76.04534 (6.7)	−76.05579 (0.14)	−76.05601
$(H_2O)_2$	−152.09870 (13.1)	−152.11909 (0.3)	−152.11957
H_2CO	−113.89717 (7.6)	−113.90892 (0.25)	−113.90932
C_2H_2	−76.84482 (1.8)	−76.84766 (0.03)	−76.84770
C_2H_4	−78.04564 (9.9)	−78.06083 (0.43)	−78.06151
C_2H_6	−79.23820 (12.0)	−79.25669 (0.37)	−79.25728
CO	−112.77502 (1.2)	−112.77682 (0.07)	−112.77694
CO_2	−187.68615 (9.3)	−187.70039 (0.31)	−187.70089
HCOOH	−188.80934 (16.6)	−188.83472 (0.70)	−188.83584
CH_3Li	−47.02843 (4.2)	−47.03497 (0.1)	−47.03519
CH_3F	−139.06451 (18.7)	−139.09268 (1.0)	−139.09437
CH_3Cl	−499.12694 (6.3)	−499.13655 (0.4)	−499.13713
CH_3OH	−115.05402 (20.0)	−115.08420 (1.07)	−115.08591
CH_3NH_2	−95.22077 (18.4)	−95.24851 (0.95)	−95.25003
CH_3CN	−131.94848 (13.9)	−131.96943 (0.1)	−131.97068
$HCONH_2$	−168.97399 (14.5)	−168.99614 (0.5)	−168.99702
$HCOCH_3$	−152.94509 (16.0)	−152.96944 (0.8)	−152.97066
C_2H_3Cl	−536.96618 (14.1)	−536.98634 (1.4)	−536.98861
$C_2H_3CH_3$	−117.08106 (19.7)	−117.11072 (1.1)	−117.11254
HCOCl	−572.83636 (9.6)	−572.85123 (0.3)	−572.85165
ZrH_4	−48.61707 (3.9)	−48.62285 (0.3)	−48.62327

[a] Relative energies compared to HF are given in parentheses (kcal/mol)

3 Applications

3.1 The relative stability of propene versus cyclopropane

In order to test the reliability of the method in problematic cases, the energy difference between two isomers of C_3H_6 has been calculated at various levels of theory. Since the two-atom bond localization constraint will be different between σ- and π-bonds, it is expected that the relative energy difference between cyclopropane and propene at the ELMO and HF levels will be different. The geometry of both isomers are optimized at the HF level in a triple-ζ + polarization basis set [20] (CH_3 is constrained to possess a C_3 axis) and the two-atom bond orbitals were optimized at the ELMO level of theory. The ELMO(JC) theory was then used to obtain the delocalization effects. The results are presented in Table 7. The differences between the ELMO and HF energies are small, 19.0 and 24.8 kcal/mol for propene and cyclopropane, respectively. However, the relative stability of propene versus cyclopropane is 5.8 kcal/mol larger than the HF value. This large difference can be explained by the different localization energy constraint in the σ and π two-atom bonds. Allowing the delocalization in the treatment of ELMO(JC) gives a relative energy between the two isomers of 8.1 kcal/mol compared to 8.0 kcal/mol at the HF level. The energy differences between ELMO(JC) and HF for both isomers reaches chemical precision.

3.2 The π-bond orbital delocalization

The use of a two-atom bond is problematic in cases of π-orbital delocalization of π-conjugation. However, it will be shown in several examples that these delocalization effects may be well described at the ELMO level. However, in order to account for the reactivity of a

Table 7. Comparison of ELMO, ELMO(JC) and HF total energies[a, b] for cyclopropane and propene

	ELMO	ELMO(JC)	ELMO + 1 iter SCF	HF
Cyclopropane	−117.05700 (24.8)	−117.09480 (1.1)	−117.09495 (1.0)	−117.09651
Propene	−117.07886 (19.0)	−117.10765 (1.0)	−117.10773 (0.9)	−117.10919
stability (kcal/mol)	13.72	8.06	8.02	7.96

[a] Total energies are given in atomic units
[b] Relative energies compared to HF are given in parentheses (kcal/mol)

Table 8. Comparison of ELMO, ELMO(JC) and HF for different π delocalization patterns in propene, propenal, butadiene and benzene

	Propene	Propenal	Butadiene	Benzene
ELMO[a]	−117.08106 (19.7)	−190.78721 (25.5)	−154.92386 (30.4)	−230.57819 (121)
ELMO[b]	−117.08464 (17.5)	−190.80070 (17.0)	−154.94043 (20.0)	−230.71481 (35.6)
ELMO[c]		−190.80160 (16.5)	−154.94138 (19.4)	−230.72228 (30.9)
ELMO(JC)[a]	−117.11072 (1.1)	−190.82432 (2.2)	−154.96804 (2.7)	−230.72223 (30.9)
ELMO(JC)[b]	−117.11128 (0.8)	−190.82574 (1.3)	−154.97051 (1.1)	−230.76519 (4.0)
ELMO(JC)[c]		−190.82616 (1.1)	−154.97099 (0.8)	−230.76652 (3.0)
HF	−117.11254 (0.0)	−190.82785 (0.0)	−154.97338 (0.0)	−230.77157 (0.0)

[a] Two-atom bond localization
[b] Delocalization to first neighbors heavy atoms
[c] Full π-delocalization to heavy atoms. Relative energies compared to HF are given in parentheses (kcal/mol)

π-bond in a given reaction, some additional delocalization may be necessary. In the latter case, the delocalization of the π two-atom bond to its closest heavy atom neighbors gives results extremely close to the fully delocalized π-orbitals.

In order to study the delocalization of a π-bond to the closest heavy atom neighbors, the propene molecule is examined in detail. The geometry is fully optimized at the HF level in a triple-ζ + polarization basis set [20]. The first calculation is performed with pure two-atom bond orbitals. This results in an ELMO energy higher than the HF energy by 19.7 kcal/mol, reduced to 1.1 kcal/mol by including the delocalization effects. A second calculation is performed in which the π two-atom bond is optimized with a delocalization over the CGTOs of the methyl carbon atom. The optimized ELMO energy is then higher than the HF energy by 17.5 kcal/mol, or 0.8 kcal/mol after treatment of the delocalization effects. Though the delocalization of the π-bond at the ELMO level gives a 2.2 kcal/mol stabilization in energy, the JC treatment of the delocalization effects (σ and π) was able to take these effects into account accurately, as the difference is reduced to only 0.3 kcal/mol. However, in order to properly describe a nucleophilic reaction in the β-position of a double bond (as in ketones), the delocalization of the π-bond needs to be addressed.

Two other systems are studied to show the relative importance of the delocalization effects of two conjugated π-bonds: butadiene and propenal. The geometry is fully optimized at the HF level in a triple-ζ+polarization basis set [20] and several ELMO calculations are performed with different localization patterns for the π-bonds. The results are summarized in Table 8. The JC takes into account a large amount of the delocalization effects of two-atom bonds. The ELMO(JC) energy is above the HF energy by 2.2 and 2.7 kcal/mol for propenal and butadiene, respectively. Allowing for a possible delocalization of the π-bonds towards the closest heavy atoms reduces the relative difference of the ELMO energy versus the HF energy by 8.5 and 10.4 kcal/mol for propenal and butadiene, respectively. The σ and the remaining π-delocalization effects are treated efficiently by the JC and the ELMO(JC) energy is now higher than the HF energy by only 1.3 and 1.1 kcal/mol. A full delocalization of the π-bonds only reduces the ELMO energy by approximately 0.5 kcal/mol and the ELMO (JC) is only reduced by 0.1 kcal/mol for both molecules. Although the Jacobi treatment is efficient in evaluating the delocalization effects these examples show the importance of the optimization of these delocalized bonds and particularly the importance of local effects of the closest neighbor atoms.

The aromaticity of benzene is considered next. Using two-atom bonds, only one Kekule structure can be represented and thus it is expected that delocalization effects are extremely important in this case. At the fully optimized HF geometry in a triple-ζ + polarization basis set [20], a two-atom bond ELMO calculation, results in a total energy higher than the HF energy by 121.3 kcal/mol. The treatment of the delocalization effects reduces this difference to 30.9 kcal/mol and clearly demonstrates the inability of π two-atom bonds to describe the electronic structure of benzene. When the localization pattern of a π-bond is extended to the two closest neighbors, the optimized ELMO energy is then higher than the HF energy by 35.6 kcal/mol without JC of the σ (and remaining π) delocalization effects, and 4.0 kcal/mol with JC. Usually fully delocalized π-orbitals, the respective numbers are reduced to 30.9 and 3.2 kcal/mol. The treatment of the delocalization effect is accurate enough to treat the extension of a bond localization to the second neighbor atoms, but the delocalization to the first neighbors need to be optimized at the ELMO level.

It should be obvious that the delocalization of a bond to its neighbors increases the number of variables to be optimized. But in the case of π-orbitals, described with p_π-CGTOs, the number of variables for delocalizing an orbital to its closest neighbors is lower than the number of variables in a σ-orbital. Thus delocalizing the π-orbitals to the closest heavy atom neighbors does not represent any particularly burdensome optimization difficulty while insuring the reliability of the result. Systems involving transition metal bonding to the π-systems of conjugated molecule (benzene, cyclopentadienyl) will be presented in future work.

3.3 The two-atom bond orbital transferability

One of the key features of the ELMO method is the power to generate and take advantage of extremely accurate guess orbitals and thus enhance the convergence properties of the gradient method by reducing the number of iterations to converge. As a test for

Table 9. Correspondence guess orbital small system

System	Orbital[a]
CH_4	σC_3H_1, σC_3H_2, σC_3H_3
$H_2C{=}CH_2$	σC_2H_4, σC_1C_2, πC_1C_2
H_2CO	σC_4H_5, nbsO, nbpO, σC_4O, πC_4O
$H_2C{=}CHCl$	nbsCl, nbpCl, σC_1Cl, πC_1Cl
CH_3CHO	σC_2C_4
$CH_3CH{=}CH_2$	σC_1C_3

[a] The atom numbering is given in Fig. 6

transferability of two-atom bond orbitals a relatively small molecule (so that an HF calculation can be performed in the same triple-ζ+ polarization basis set [20]) is chosen: $CH_3ClC{=}CHCHO$ (4-chlorobut-2-enal). The guess two-atom bonds were taken from the optimized two-atom bonds of small systems, rotated and normalized to unity. As a first approximation, the two-atom bond orbital basis set expansion is considered independent of the bond distance and calculated at the small system's optimized geometry. One can imagine a future refinement by making a database of two-atom bond orbitals accessed both by the atom bond type and bond distance. The small systems used to extract the guess orbitals are given in Table 9, and intentionally neither represent the best choice of fragment partition nor include any π-delocalization effects. The ELMO optimization procedure lowers the total energy by 29.8 kcal/mol and the treatment of the delocalization effects lowers it by 29.7 kcal/mol. The final ELMO(JC) total energy is higher than the HF energy by only 1.4 kcal/mol (Table 10).

3.4 The metal CO bond

The bonding of carbon monoxide (CO) to a transition metal atom is an interesting example of σ-donation from ligand to metal and π-backdonation from metal to ligand. The H_2ZrCO model system has been studied in order to determine necessary conditions on the localization patterns. The electronic structure of isolated CO can be accurately described at the ELMO level in terms of optimized localized bond orbitals with the following orbitals (the atoms on which an orbital is localized are given in parenthesis): 1*s*C(C), 1*s*O(O), non-bonding hybrid orbital O(O), σ_{CO}(CO), π_{xCO}(CO), π_{yCO}(CO) and non-bonding hybrid orbital C(C). The ELMO energy is higher than the HF energy by only 1.2 kcal/mol at the ELMO level and less than 0.1 kcal/mol at the JC level. Thus, the treatment of the σ-delocalization effects in the CO group is accurate.

The geometry of H_2ZrCO was optimized at the HF level. An effective core potential was used for the Zr metal atom associated with a $(3s3p3d)$ basis set [33] and a triple-ζ basis [20] set on all other atoms. The optimized geometry is shown in Fig. 3. Several ELMO calculations were performed with different localization patterns. In all the calculations the 1*s*C(C), 1*s*O(O), non-bonding hybrid orbital O(O), σ_{CO}(CO), π_{xCO}(CO) and π_{yCO}(CO) orbitals were optimized locally in the CO fragment. The

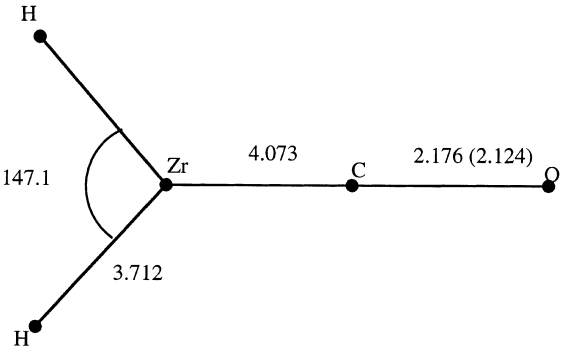

Fig. 3. Optimized HF geometry of H_2ZrCO. All bond distances in a.u. and angles in degrees. The equilibrium bond distance of CO (optimized at the HF level in the triple-ζ basis set [17] is given in parentheses)

two Zr—H bond orbitals also were optimized as two-atom bonds in all calculations. The possible delocalization effects, examined here, are through σ-donating and π-backdonating orbitals. These orbitals were localized among the Zr, C and O atoms. The results are summarized in Table 11. Although the effect of restricting the localization of the σ_{Zr-CO} to the Zr and C atoms is very small (0.08 kcal/mol), this restriction on the π_{Zr-CO} has a larger effect (2.68 kcal/mol). However, the treatment of delocalization seems to be able to compensate for the localization restriction. Since the number of variables necessary to optimize a "delocalized" π_{Zr-CO} orbital is relatively small, a three-atom bond can be used to more accurately describe the π-backdonation effects. Contour plots of the σ_{Zr-CO} and π_{Zr-CO}, delocalized over Zr, C and Zr, C, O respectively, are given in Figs. 4 and 5. In this test case, the HF/ELMO(JC) energy difference is less than 0.5 kcal/mol.

Table 10. Total ELMO, ELMO(JC) and HF energies[a] for $CH_3ClC{=}CHCHO$

	Total energy (a.u.)
ELMO (1st iteration = HF energies with guess orbitals)	−688.69965 (60.9)
ELMO (convergence)	−688.74720 (31.0)
ELMO (JC)	−688.79448 (1.40)
HF	−688.79668 (0.00)

[a] Relative energies compared to HF are given in parentheses (kcal/mol)

Table 11. Total ELMO, ELMO(JC) and HF energies for H_2ZrCO with different localization patterns

	ELMO (a.u.)	ELMO (JC) (a.u.)
σ (Zr, C, O); π (Zr, C, O)	−160.15043 (3.9)	−160.15632 (0.2)
σ (Zr, C); π (Zr, C, O)	−160.15031 (4.0)	−160.15632 (0.2)
σ (Zr, C, O); π (Zr, C)	−160.14617 (6.6)	−160.15611 (0.3)

[a] Relative energies compared to HF are given in parentheses (kcal/mol). E_{HF} = −160.15664 a.u.

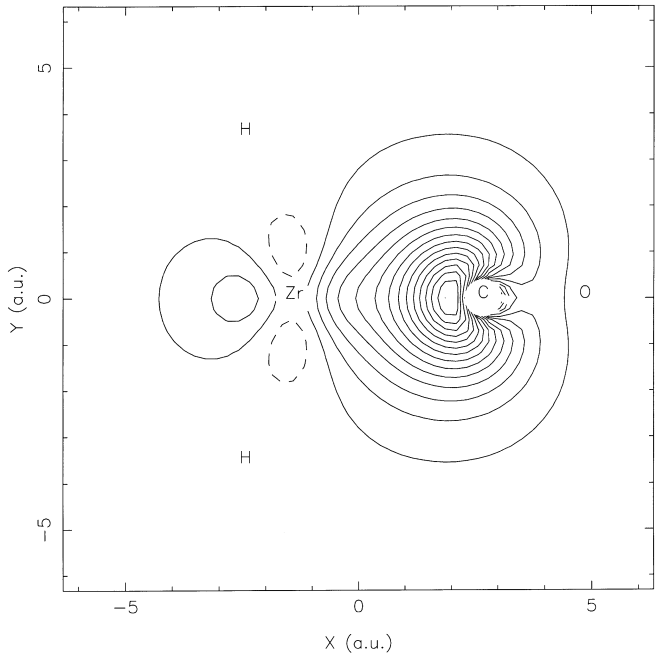

Fig. 4. Contour plot of the σ_{ZrC} two-atom bond of H_2ZrCO in the plane of the molecule

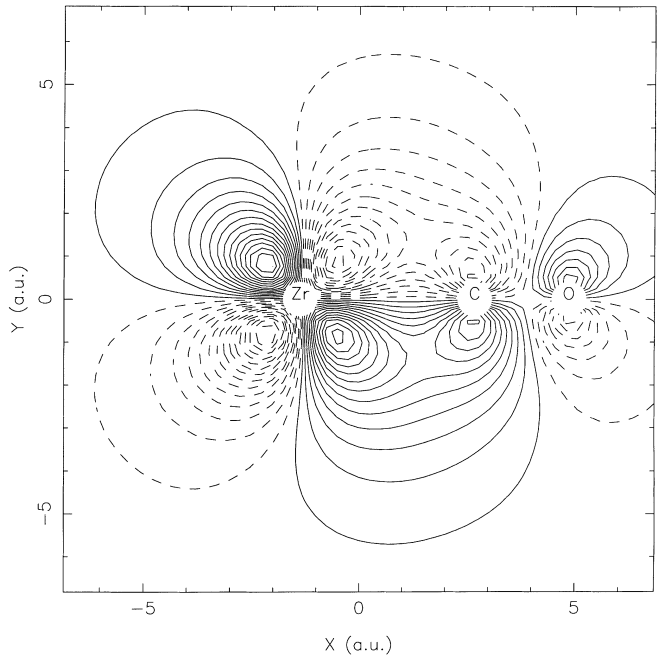

Fig. 5. Contour plot of the π_{ZrCO} orbital of H_2ZrCO in a plane orthogonal to the molecule containing the three atoms

4 Conclusion

The ELMO method encompasses several important features of MO optimization codes such as low-order scaling, efficient convergence and ability to take full advantage of a high quality set of initial orbitals. For this method, an efficient variable metric or quasi Newton-Raphson algorithm has been derived in a localized non-orthogonal formalism. This algorithm requires a good approximate Hessian matrix to reach quadratic convergence and the ELMO method is based on an efficient derivation of an approximate Hessian matrix through the use of the Fock matrix only. The data structure of the localized orbitals allows the Hessian to be set up in diagonal-blocked form, ignoring the contribution of the orbital coupling elements which are introduced during the course of the iterations by a DFP or BFGS updating technique. An orbital optimization iteration in ELMO consists of building one Fock matrix as in standard HF methods and updated orbitals are calculated with approximately n operations (n^2 operations are required to evaluate the gradient of the energy versus the variables). The method is free from any large n^3 processes such as Fock matrix diagonalization, overlap matrix diagonalization, matrix multiplication or orbital orthogonalization. The delocalization effects (contributing to an energy stabilization of 10 to 30 kcal/mol) are introduced in ELMO(JC) by a treatment using the Fock matrix built in the last iteration. The delocalized orbitals allow calculation of the total energy while requiring only a small number of additional two-electron integrals.

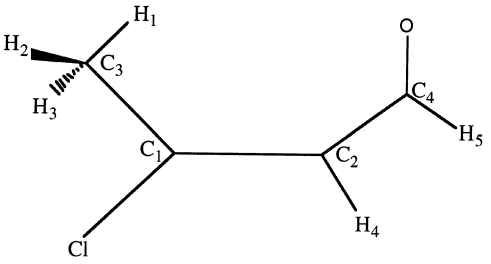

Fig. 6. Atom numbering in $CH_3ClC\text{-}CHCHO$

The Fock matrix computation can take full advantage of the currently developed fast multipole methods [7] and thus reach linear scaling. Since the Fock matrix computation can also be obtained efficiently over several hundreds of processors [34], the ELMO method can be fully optimized for parallel architectures using the N-separability of the Newton-Raphson procedure. The treatment of the delocalization effects in the ELMO(JC) can also use the efficiency of the inherent parallelism of the Jacobi algorithm. Research with these perspectives is currently under development and will be presented in the near future.

Appendix

Intrinsic zero eigenvalues of the Hessian matrix

The energy formula (Eq. 11) is exact using both normalized or unnormalized non-orthogonal orbitals. Thus a step vector $\delta\mathbf{X}$ proportional to $\mathbf{X}$ would only change the norm of the orbitals without any effect on the total energy. Since this artifact is not explicitly taken into account in the energy minimization process, each block of the exact diagonal-blocked Hessian matrix has an intrinsic eigenvalue equal to zero corresponding to an eigenvector equal to the molecular orbital development. However, it can be demonstrated that this vector is orthogonal to the gradient vector (see demonstration below). A block of the exact diagonal-blocked Hessian matrix may also have other intrinsic zero eigenvalues corresponding to the fact that the energy is invariant to the mixing of certain

molecular orbitals. If the mixing between several orbitals is obtained while maintaining the localization pattern, the blocked Hessian matrix has some additional zero eigenvalues. It can be shown that the corresponding eigenvectors are orthogonal to the gradient vector. This situation occurs when core or non-bonding orbitals (developed only on the CGTOs of one atom) are present. This problem can be encountered in a more general way if the localization of orbitals is extended to a larger number of atoms and reaches its maximum amplitude upon using fully delocalized orbitals.

Theorem: The gradient vector corresponding to a given orbital k is orthogonal to any occupied molecular orbital 1.

To demonstrate the theorem, the following quantity is evaluated:

$$\sum_a \frac{dE}{dC_{ak}} C_{al} = \sum_{p,q} F_{pq} \sum_a \frac{d\rho_{pq}}{dC_{ak}} C_{al} \tag{A1}$$

$$\begin{aligned}\sum_a \frac{d\rho_{pq}}{dC_{ak}} C_{al} &= \sum_a \left(\delta^a_p A_{kq} + \delta^a_q A_{kp} - A_{kp}A^*_{aq} - A_{kq}A^*_{ap}\right) C_{al} \\ &= A_{kq}C_{pl} + A_{kp}C_{ql} - A_{kp}\sum_a A^*_{aq}C_{al} \\ &\quad - A_{kq}\sum_a A^*_{ap}C_{al} \\ &= A_{kq}C_{pl} + A_{kp}C_{ql} - A_{kp}({}^tA^*C)_{ql} \\ &\quad - A_{kq}({}^tA^*C)_{pl} \ .\end{aligned} \tag{A2}$$

The following expressions allow the calculation of $({}^t\mathbf{A}^*\mathbf{C})_{pl}$:

$$\begin{aligned}&\mathbf{A}^* = {}^t\mathbf{B}\mathbf{A} \\ &\mathbf{B}^* = {}^t\mathbf{B}\mathbf{S}^{-1} \\ &\mathbf{A} = \mathbf{S}^{-1\,t}\mathbf{C} \\ &\mathbf{B} = {}^t\mathbf{C}\sigma \\ &\mathbf{S} = {}^t\mathbf{C}\sigma\mathbf{C} \ ,\end{aligned} \tag{A3}$$

$$\begin{aligned}({}^t\mathbf{A}^*\mathbf{C})_{pl} &= ({}^t\mathbf{A}\mathbf{B}\mathbf{C})_{pl} = (\mathbf{C}\mathbf{S}^{-1\,t}\mathbf{C}\sigma\mathbf{C})_{pl} \\ &= (\mathbf{C}\mathbf{S}^{-1}\mathbf{S})_{pl} = \mathbf{C}_{pl} \ .\end{aligned} \tag{A4}$$

Thus:

$$\sum_a \frac{d\rho_{pq}}{dC_{ak}} C_{al} = 0, \ \forall k, l$$

$$\sum_a \frac{dE}{dC_{ak}} C_{al} = \sum_{p,q} F_{pq} 0 = 0 \ . \tag{A5}$$

Theorem: The block corresponding to a given orbital k in the approximate blocked Hessian matrix has some zero eigenvalues associated with eigenvectors corresponding to an occupied molecular orbital l.

To demonstrate the theorem, the following quantity is evaluated:

$$(H^k C)_{al} = \sum_b \frac{d^2E}{dC_{ak}dC_{bk}} C_{bl} = \sum_{p,q} F_{pq} \sum_b \left(\frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}}\right)^* C_{bl}$$

$$\begin{aligned}&\sum_b \left(\frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}}\right)^* C_{bl} \\ &= S^{-1}_{kk}\left(\delta^a_p C_{ql} + \delta^a_q C_{pl} - \delta^a_p \sum_b A^*_{bq}C_{bl} - \delta^a_q \sum_b A^*_{bp}C_{bl}\right) \\ &\quad + S^{-1}_{kk}\left(-A^*_{aq}C_{pl} - A^*_{aq}C_{ql} + A^*_{aq}\sum_b A^*_{bp}C_{bl}\right. \\ &\quad \left. + A^*_{ap}\sum_b A^*_{bq}C_{bl}\right) + 2A_{kp}A_{kq}\left(\sum_{i,j=1}^{N} B_{ia}S^{-1}_{ij}\sum_b B_{jb}C_{bl}\right. \\ &\quad \left. - \sum_b \sigma_{ab}C_{bl}\right) \ .\end{aligned} \tag{A6}$$

Using Eq. (A4), the following expression can be simplified to:

$$\sum_b \left(\frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}}\right)^* C_{bl} = 2A_{kp}A_{kq}\left[({}^tBS^{-1}BC)_{al} - (\sigma C)_{al}\right] \ . \tag{A7}$$

Since $({}^t\mathbf{B}\mathbf{S}^{-1}\mathbf{B}\mathbf{C})_{al} = (\sigma\mathbf{C}\mathbf{S}^{-1\,t}\mathbf{C}\sigma\mathbf{C})_{al} = (\sigma\mathbf{C}\mathbf{S}^{-1}\mathbf{S})_{al} = (\sigma\mathbf{C})_{al}$, the Eq. (A7) is equal to 0 for all couples (a, l), i.e., for all occupied molecular orbitals. Thus, any molecular orbital l that has a localization pattern included in the localization pattern of orbital k is associated with a zero eigenvalue of the approximate block Hessian relative to orbital k.

N.B.

$$\begin{aligned}\sum_b \frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}} C_{bl} &= \sum_b \left(\frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}}\right)^* C_{bl} \\ &\quad + \left(-\delta^a_p A_{kq} - \delta^a_q A_{kp} + A_{kp}A^*_{aq} + A_{kq}A^*_{ap}\right)\sum_b B^*_{bk}C_{bl} \\ &\quad + B^*_{ak}\left(-A_{kq}C_{pl} - A_{kp}C_{ql} + A_{kp}\sum_b A^*_{bq}C_{bl}\right. \\ &\quad \left. + A_{kq}\sum_b A^*_{bp}C_{bl}\right) \ .\end{aligned} \tag{A8}$$

Using Eq. (A4), the previous equation can be simplified:

$$\begin{aligned}&\sum_b \frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}} C_{bl} \\ &\quad = \left(-\delta^a_p A_{kq} - \delta^a_q A_{kp} + A_{kp}A^*_{aq} + A_{kq}A^*_{ap}\right)\delta^l_k \ .\end{aligned} \tag{A9}$$

Since

$$\begin{aligned}\sum_b B^*_{bk}C_{bl} &= ({}^t\mathbf{B}^*\mathbf{C})_{kl} = (\mathbf{S}^{-1}\mathbf{B}\mathbf{C})_{kl} \\ &= (\mathbf{S}^{-1\,t}\mathbf{C}\sigma\mathbf{C})_{kl} = (\mathbf{S}^{-1}\mathbf{S})_{pl} = \delta^l_k \ .\end{aligned} \tag{A10}$$

Thus, $\sum_{p,q} F_{pq} \sum_b \frac{d^2\rho_{pq}}{dC_{ak}dC_{bk}} C_{bk} = -\frac{dE}{dC_{ak}}$ vanishes only at the convergence.

References

1. Almlöf JE, Faegri K, Korsell K (1982) J Comput Chem 3:385
2. Almlöf JE, Taylor PR (1984) In: Dykstra C (ed) Advanced theories and computational approaches to the electronic structures of molecules (NATO ASI) Reidel, Dordrecht
3. Dyczmons V (1973) Theor Chim Acta 28:307; Ahlrichs R (1974) Theor Chim Acta 33:157
4. Haser M, Ahlrichs R (1989) J Comput Chem 10:104
5. Panas I, Almlöf JE (1992) Int J Quantum Chem 42:1073; Panas I, Almlöf JE, Feyereisen MW (1991) Int J Quantum Chem 40:797
6. Schwegler E, Challacombe M (1996) J Chem Phys 105:2726
7. White CA, Johnson BG, Gill PMW, Head-Gordon M (1994) Chem Phys Lett 230:8; Strain MC, Scuseria GE, Frisch MJ (1996) Science 271:51
8. Challacombe M, Schwegler E, Almlöf JE (1996) J Chem Phys 104:4685; White CA, Head-Gordon M (1996) J Chem Phys 104:2620
9. Rendell AP (1994) Chem Phys Lett 229:204; Shepard R (1993) Theor Chim Acta 84:343; Fischer TH, Almlöf JE (1992) J Phys Chem 96:9768
10. Couty M, Lévy B (1994) Int J Quantum Chem 52:59
11. Dixon SL, Merz KM (1996) J Chem Phys 104:6643
12. Stewart JJP (1996) Int J Quantum Chem 58:133
13. Roothaan CCJ (1951) Rev Mod Phys 23:69; Hall GG (1951) Proc R Soc A 208:328
14. Lévy B (1971) PhD thesis, no. AO 5271 Paris, France; Lévy B (1970) Int J Quantum Chem 4:297; Lévy B (1973) Chem Phys Lett 18:59; Bacskay GB (1981) Chem Phys 61:385; Bacskay GB (1981) Chem Phys 65:383

15. Cooper DL, Raimondi M, Gerratt J (1987) Adv Chem Phys 69:319; Gerratt J, Cooper DL, Raimondi M (1990) In: Klein DJ, Trinojstic N (eds) The spin-coupled valence bond theory of molecular electronic structure. Elsevier, Amsterdam; Van Lenthe JH, Balint-Kurti GG (1983) J Chem Phys 78:5699
16. (a) Saunders VR, Hillier IH (1974) Mol Phys 28:819; (b) Pulay P (1982) J Comp Chem 3:556
17. Paldus J (1990) In: Carbo R, Klobukowski M (eds) Self-consistent field theory and applications. Elsevier, Amsterdam p 1 and references therein
18. Pollard WT, Friesner RA (1993) J Chem Phys 99:6742
19. Lowdin PO (1955) Phys Rev 97 6:1474
20. Dunning TH (1971) J Chem Phys 55:716; McLean AD, Chandler GS (1980) J Chem Phys 52:5639
21. Boys SF (1966) In: Lowdin PO (ed) Quantum science of atoms, molecules and solids. Academic Press, New York, p 253
22. Chambaud G, Gerard-Ain M, Kassab E, Levy B, Pernot P (1984) Chem Phys 90:271
23. Stoll H, Wagenblast G, Preuss H (1980) Theor Chim Acta 57:169; Smits GF, Altona C (1985) Theor Chim Acta 67:461
24. Stechel EB, Williams AR, Feibelman PJ (1994) Phys Rev B 49:10088
25. Verbeek J (1990) PhD thesis, Utrecht, The Netherlands
26. Couty M (1993) PhD thesis no. 2848, Paris XI, France
27. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1989) Numerical recipes in Fortran, 2nd edn., Cambridge, University Press
28. Jørgensen P, Swanstrom P, Yeager D (1983) J Chem Phys 78:347
29. Geometries were optimized using GAMESS_UK. Guest MF, Kendrick J, van Lenthe JH, Schoeffel K, Sherwood P (1995) Computing for Science, Ltd., Daresbury Laboratory, Warrington, U.K.
30. Chavy C (1991) Ph D Thesis. Paris XI, France
31. Couty M, Hall MB, J Phys Chem (submitted)
32. Golub GH, Van Loan CF (1989) Matrix computations, 2nd edn. Johns Hopkins, University Press
33. Hurley MM, Fernandez Pacios L, Christiansen PA, Ross RB, Ermler WC (1986) J Chem Phys 84:6840; LaJohn LA, Christiansen PA, Ross RB, Atashroo T, Ermler WC (1987) J Chem Phys 87:2812; Ross RB, Powers JM, Atashroo T, Ermler WC, LaJohn LA, Christiansen PA (1990) J Chem Phys 93:6654; Couty M, Hall MB (1996) J Comput Chem 17:1359
34. Feyereisen M, Kendall RA (1993) Theor Chim Acta 84:289