Theoret. Chim. Acta (Berl.) 38, 211—221 (1975)
© by Springer-Verlag 1975

Direct Minimization of the Energy Functional in LCAO-MO Density Matrix Formalism

I. Closed and Open Shell Systems

Stefano Polezzo
Centro del C.N.R. "Relazioni fra struttura e reattività chimica", Milano, Italy

Received January 15, 1975 / March 10, 1975

A direct method of minimization of the energy expression for closed and open shell systems in LCAO-MO density matrix formalism is presented. The method makes use of a unitary transformation acting directly on the density matrices. Expressions of the gradient and second energy derivatives are worked out. Some preliminary calculations to test the rate of minimization using a variable metric method have been made on H_2S and SO molecules and have given satisfactory results.

Key words: Direct energy minimization – Density matrix LCAO-MO-SCF

1. Introduction

Two main approaches can be distinguished to solve equations of the SCF theory. The former originates from the classical work of Hartree and Fock and consists in solving iteratively a system of variational equations as a pseudo-eigenvalue problem of a one-electron hamiltonian. The latter is based on a direct minimization of the energy expression. Among the approaches of this kind the one proposed by McWeeny [1] approximates the energy surface near a starting point by a linear variation of density matrices, made in such a way to satisfy the constraints, with a subsequent descent along the negative gradient direction.

Recently, following the work of Fletcher [2], interest has been revived in direct methods of minimizing the energy with respect to orbital linear coefficients (or to other non-linear parameters such as nuclear positions and orbital exponents) using one of the more modern conjugate-direction technique [3]. The basic idea of this approach consists in the introduction of a matrix to incorporate and remove the constraints and of a matrix of unconstrained variables, with respect to which the minimization is performed. A drawback of this procedure seems to be that, due to the presence of redundant variables, the Hessian matrix is not positive-definite, as it should be, but singular [4].

An alternative direct minimization approach, which may be traced out to the paper by Raffenetti and Ruedenberg [5] (see also [6]), is based on the use of the constraints to remove variables from the problem. An expression of the energy in terms of independent variables can be achieved through a parametrization of an orthogonal matrix, which allows the maintenance of the orthogonality (or idempotency and exclusivity in the density matrix formalism) during all the iterative steps and so the iterative minimum seeking procedure is carried out directly on the energy surface.

In the line of the approach now mentioned, we propose here to use an orthogonal $m \times m$ matrix of the form (known as Cayley's formula)

$$\boldsymbol{U}(\boldsymbol{X}) = -\mathbf{1} + 2(\mathbf{1} + \boldsymbol{X} - \tilde{\boldsymbol{X}})^{-1}\,, \tag{1}$$

depending on the $m(m-1)/2$ independent variables given by the elements of the skew-symmetric matrix $\boldsymbol{S} = \boldsymbol{X} - \tilde{\boldsymbol{X}}$, $\boldsymbol{X}$ being a completely arbitrary (real) matrix. Introducing, for the sake of brevity, the definition

$$\boldsymbol{P} = \mathbf{1} + \boldsymbol{X} - \tilde{\boldsymbol{X}} = \mathbf{1} + \boldsymbol{S}\,, \tag{2}$$

it is worth noting that the matrix $\boldsymbol{P}$ is never singular, due to the fact that all the non-zero characteristic roots of a real skew-symmetric matrix are imaginary (and hence unity cannot be one of these roots). The unitarity of $\boldsymbol{U}$, that is

$$\boldsymbol{U}\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{U}}\boldsymbol{U} = \mathbf{1}\,,$$

is immediately shown observing that the matrices $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}$ commute, that is $\boldsymbol{P}\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{P}}\boldsymbol{P}$.

The proposal is to employ this orthogonal matrix to transform the m dimensional atomic basis set, assumed to be orthonormal: in this way a map is made of an orthonormal vector set into another orthonormal set, and the number of independent variables needed to do a transformation of this kind is just $m(m-1)/2$.

In this note we deal with a rather general energy expression for many-shells systems in terms of density matrices; the first and second energy derivatives with respect to the independent variables are worked out and applied to the direct minimization of the energy functional. In a subsequent note the method will be extended to a more general energy expression including configuration interaction.

2. Theory

A straightforward extension of the original McWeeny theory [1] to states requiring more than one determinant for configurations involving many shells leads to an energy expression which may be written as follows

$$E = \sum_{i=1}^{n_s} v_i \operatorname{tr}(\boldsymbol{f}\boldsymbol{R}_i) + \sum_{i \leq j}^{n_s} v_{ij} \operatorname{tr}(\boldsymbol{R}_i \boldsymbol{G}_{ij}(\boldsymbol{R}_j))\,. \tag{3}$$

In this expression i and j denote the shells, n_s is the number of shells and v_i their fractional occupation number ($0 < v_i \leqq 2$); $\boldsymbol{R}_i$ is the density matrix of the i^{th} shell and $v_{ij} = v_i v_j/(1 + \delta_{ij})$. According to its definition, v_i may assume also fractional values to allow all the orbitals in a degenerate set to have the same occupation number. $\boldsymbol{f}$ denotes the matrix of one-electron integrals and $\boldsymbol{G}_{ij}(\boldsymbol{R}_j)$ the usual electron interaction matrix of elements

$$[\boldsymbol{G}_{ij}(\boldsymbol{R}_j)]_{rs} = \sum_{t,u}^{m} R_{jtu}\, g_{ru,st}^{ij}\,, \tag{4}$$

where the notation

$$g_{ru,st}^{ij} = a_{ij}\langle ru|st\rangle - b_{ij}\langle ru|ts\rangle \tag{5}$$

has been introduced for a combination of two-electron integrals on an atomic orbital basis of dimension m. The coefficients a_{ij} and b_{ij} are constants that couple the shells i and j; their values are determined by writing down explicitly the energy expression that arises from the several Slater determinant wave function, associated with the state of the system (appropriate spin and spatial symmetry). Due to the orthonormality requirement among the molecular orbitals, the hermitian matrices $\boldsymbol{R}_i$ must be idempotent and mutually exclusive, that is, in an orthonormal basis,

$$\boldsymbol{R}_i^\dagger = \boldsymbol{R}_i\,, \qquad \boldsymbol{R}_i^2 = \boldsymbol{R}_i\,, \qquad \boldsymbol{R}_i \boldsymbol{R}_j = \delta_{ij} \boldsymbol{R}_i \qquad (i,j = 1, \dots n_s) \tag{6}$$

with traces having the integer values

$$\operatorname{tr} \boldsymbol{R}_i = n_i \qquad (i = 1, \dots n_s)\,, \tag{7}$$

corresponding to the n_i orbitals with occupation number v_i forming the shell i. Each density matrix $\boldsymbol{R}_i$ represents a projection operator defining a n_i-dimensional subspace of the m-dimensional space spanned by the atomic orbitals.

The energy expression (3) has to be minimized with respect to the linear coefficients of molecular orbitals under the constraints (6) and (7). Introducing an orthogonal matrix $\boldsymbol{U}(\boldsymbol{X})$ and starting with density matrices $\boldsymbol{R}_i$ satisfying (6) and (7), the transformed matrices

$$\bar{\boldsymbol{R}}_i = \boldsymbol{U}(\boldsymbol{X})\, \boldsymbol{R}_i\, \tilde{\boldsymbol{U}}(\boldsymbol{X}) \tag{8}$$

continue to be symmetric and to satisfy the conditions (6) and (7). Of course, in the presence of symmetry the matrix $\boldsymbol{U}(\boldsymbol{X})$ has the same block structure of the density matrices $\boldsymbol{R}_i$ and as a consequence the number of independent variables is accordingly decreased. Inserting the $\bar{\boldsymbol{R}}_i$, the expression (3) of the energy becomes

$$E(\boldsymbol{X}) = \sum_{i=1}^{n_s} v_i \operatorname{tr}(\boldsymbol{f} \bar{\boldsymbol{R}}_i) + \sum_{i \leq j}^{n_s} v_{ij} \operatorname{tr}(\bar{\boldsymbol{R}}_i \boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_j))\,. \tag{9}$$

At this point a brief digression is in order. It is worth noting that the matrix transformation (8) is equivalent to perform, by means of the matrix $\boldsymbol{U}$, an orthogonal transformation on the atomic orbital basis, which, as said, is assumed to form an orthonormal set: the transformation induced on the density matrices is just given by (8), and the expression of the energy takes the form (9). However, from a more general point of view, it is easy to show that, given a density matrix $\boldsymbol{R}_i$, any other density matrix having the same trace can always be written in the form (8): for the sake of completeness, a proof is reported in the Appendix 1 (Part A).

Taking for $\boldsymbol{U}(\boldsymbol{X})$ the formula (1), the expression (9) can be differentiated with respect to the arbitrary variables $\boldsymbol{X}$ in order to obtain the gradient. In this context the initial matrices $\boldsymbol{R}_i$ are regarded as fixed: they are the initial point on the energy surface from which to start the transformation (8).

From the

$$\delta E(\boldsymbol{X}) = 2 \sum_{i=1}^{n_s} v_i \operatorname{tr}(\boldsymbol{R}_i \tilde{\boldsymbol{U}} \boldsymbol{f} \delta \boldsymbol{U}) + 2 \sum_{i \leq j}^{n_s} v_{ij} [\operatorname{tr}(\boldsymbol{R}_i \tilde{\boldsymbol{U}} \boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_j) \delta \boldsymbol{U}) + \operatorname{tr}(\boldsymbol{R}_j \tilde{\boldsymbol{U}} \boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_i) \delta \boldsymbol{U})] \tag{10}$$

and from the first order variation of $\boldsymbol{U}$

$$\delta \boldsymbol{U} = 2\delta \boldsymbol{P}^{-1} = -2\boldsymbol{P}^{-1}\delta \boldsymbol{S}\boldsymbol{P}^{-1} = 2\boldsymbol{P}^{-1}\delta \tilde{X}\boldsymbol{P}^{-1} - 2\boldsymbol{P}^{-1}\delta X\boldsymbol{P}^{-1}, \tag{11}$$

one obtains

$$\begin{aligned}\delta E(X) = {} & 4\sum_{i=1}^{n_s} v_i[\mathrm{tr}(\boldsymbol{P}^{-1}\boldsymbol{R}_i\tilde{\boldsymbol{U}}\boldsymbol{f}\boldsymbol{P}^{-1}\delta\tilde{X}) - \mathrm{tr}(\boldsymbol{P}^{-1}\boldsymbol{R}_i\tilde{\boldsymbol{U}}\boldsymbol{f}\boldsymbol{P}^{-1}\delta X)] \\ & + 4\sum_{i\leqq j}^{n_s} v_{ij}[\mathrm{tr}(\boldsymbol{P}^{-1}\boldsymbol{R}_i\tilde{\boldsymbol{U}}\boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_j)\boldsymbol{P}^{-1}\delta\tilde{X}) - \mathrm{tr}(\boldsymbol{P}^{-1}\boldsymbol{R}_i\tilde{\boldsymbol{U}}\boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_j)\boldsymbol{P}^{-1}\delta X) \\ & + \mathrm{tr}(\boldsymbol{P}^{-1}\boldsymbol{R}_j\tilde{\boldsymbol{U}}\boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_i)\boldsymbol{P}^{-1}\delta\tilde{X}) - \mathrm{tr}(\boldsymbol{P}^{-1}\boldsymbol{R}_j\tilde{\boldsymbol{U}}\boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_i)\boldsymbol{P}^{-1}\delta X)].\end{aligned} \tag{12}$$

Observing that $\boldsymbol{P}^{-1} = \tilde{\boldsymbol{P}}^{-1}\boldsymbol{U}$ and introducing the hamiltonian

$$\boldsymbol{h}_i = v_i\left[\boldsymbol{f} + \sum_{j=1}^{n_s} v_j\boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_j)\right], \tag{13}$$

the gradient $\boldsymbol{G}_x$ with respect to the X variables is then given by the $m \times m$ skew-symmetric matrix

$$\boldsymbol{G}_x = -4\tilde{\boldsymbol{P}}^{-1}\left[\sum_{i=1}^{n_s}\boldsymbol{h}_i\bar{\boldsymbol{R}}_i - \sum_{i=1}^{n_s}\bar{\boldsymbol{R}}_i\boldsymbol{h}_i\right]\boldsymbol{P}^{-1}. \tag{14}$$

Hence the stationary conditions are given by ($\boldsymbol{G}_x = \boldsymbol{0}$)

$$\sum_{i=1}^{n_s}\boldsymbol{h}_i\bar{\boldsymbol{R}}_i = \sum_{i=1}^{n_s}\bar{\boldsymbol{R}}_i\boldsymbol{h}_i. \tag{15}$$

Due to the skew-symmetry of $\boldsymbol{G}_x$ matrix these are $m(m-1)/2$ equations which, together with the conditions (6) and (7), completely determine the solution. As shown in Appendix 1 (Part B), the (15) are equivalent to the conditions at the minimum obtained by McWeeny [7]. Second derivatives are also easily obtained. For instance, starting from the coefficients of $\delta\tilde{\boldsymbol{U}}$ under the trace operation in (10) and deriving their expression further, after a little manipulation the following expression is arrived at

$$\begin{aligned}\frac{\partial^2 E}{\partial U_{mn}\partial U_{kl}} = {} & 2\sum_{i=1}^{n_s} v_i f_{km} R_{inl} \\ & + 2\sum_{i\leqq j}^{n_s} v_{ij}\Bigg\{[\boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_j)]_{km}R_{inl} + \sum_{t,u}^{m}(\boldsymbol{R}_j\tilde{\boldsymbol{U}})_{nt}(g^{ij}_{tu,mk} + g^{ij}_{tk,mu})(\boldsymbol{U}\boldsymbol{R}_i)_{ul} \\ & + [\boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_i)]_{km}R_{jnl} + \sum_{t,u}(\boldsymbol{R}_i\tilde{\boldsymbol{U}})_{nt}(g^{ij}_{tu,mk} + g^{ij}_{tk,mu})(\boldsymbol{U}\boldsymbol{R}_j)_{ul}\Bigg\}.\end{aligned} \tag{16}$$

For small variations, as it happens near the minimum, $\boldsymbol{U}(X)$ may be approximated at first order as ($k = -2$)

$$\boldsymbol{U}(X) = \boldsymbol{1} + k(X - \tilde{X}) \tag{17}$$

and then the final expression for the second derivatives, in this approximation, is given by

$$\frac{\partial^2 E}{\partial X_{mn} \partial X_{kl}} = k^2 \left(\frac{\partial^2 E}{\partial U_{mn} \partial U_{kl}} - \frac{\partial^2 E}{\partial U_{nm} \partial U_{kl}} - \frac{\partial^2 E}{\partial U_{mn} \partial U_{lk}} + \frac{\partial^2 E}{\partial U_{nm} \partial U_{lk}} \right).$$

These are the elements (km, ln) of a supermatrix, which will be called D_{xx}. It is thought partitioned in the usual manner. The $m \times m$ matrix formed by the elements $(m = k, n = l)$, that is $\partial^2 E/\partial X_{kl}^2$, of this supermatrix will be denoted by $\boldsymbol{D}_x$ and called diagonal block for brevity; it is easily shown that $\boldsymbol{D}_x$ is a symmetric matrix with zeros along the main diagonal, as it must be.

More general but, of course, more cumbersome expressions of second derivatives may be worked out without introducing the linear approximation (17) for $\boldsymbol{U}(\boldsymbol{X})$. This expression can be obtained deriving a second time the coefficients of $\delta \tilde{\boldsymbol{X}}$ under the trace operation of (12), as it is outlined in Appendix 4, but it has not been used in the preliminary test calculations reported below.

3. The Process of Minimization

Equation (14), being the condition that $E(\boldsymbol{X})$ is stationary with respect to the independent variables $\boldsymbol{X}$, may be satisfied by direct minimization. When an energy minimum is reached all the gradient components must vanish: if so, approaching the extremum point the elements $(\boldsymbol{G}_x)_{rs}$ of the gradient matrix diminish by absolute value. Taking the modulus m_g of the gradient so defined

$$m_g = \mathrm{tr}(\boldsymbol{G}_x \tilde{\boldsymbol{G}}_x),$$

remembering that $m_g = 0$ if and only if $\boldsymbol{G}_x = \boldsymbol{0}$, the vanishing of m_g may be taken as a criterium to test if an extremum point is near.

In a preliminary way we have tested three algorithms of minimization: a conjugate gradient, a steepest descent without second derivatives and a second derivatives algorithm, within the approximations mentioned above. Of course, the algorithms just mentioned constitute the third section of a general SCF computation program of which, the first section calculates the one- and two-electrons integrals and builds the necessary matrices, and the second section performs firstly the transformation over symmetry adapted orbitals and after their orthonormalization by the Schmidt method.

Within the class of the conjugate gradient methods the choice was made to carry out a test using, mainly for its simplicity, the variable metric minimization algorithm proposed by Murtagh and Sargent [8]: for easy reference, in the Appendix 2 the necessary formulae are rewritten in the present notations. The iteration scheme for this algorithm (and for the others to be described) can be outlined as follows. Starting with some initial density matrices $\boldsymbol{R}_i$ (for which, of course, the necessary conditions of idempotency, exclusivity and of traces are satisfied), with $\boldsymbol{X} = \boldsymbol{0}$ (that is $\boldsymbol{U} = \boldsymbol{1}$) and $\boldsymbol{A}_0 = \boldsymbol{1}$, a preliminary research is performed in order to find a point $\boldsymbol{X}$ and a value of the constant α for which both the energy and m_g are decreased with respect to the starting values. A new point $\boldsymbol{X}'$ is found along the prescriptions of the algorithm chosen, the corresponding new matrices

$U' = U(X')$ and $\bar{R}'_i = U' R_i \tilde{U}'$ are computed together with the associated gradient matrix G'_x and the values of the energy E and of the modulus m_g. The iteration is then repeated up to the prescribed precision, taking every time as new point to start from X' and the corresponding gradient G'_x. The test of convergence is made on both the values of the energy decrement ΔE and of m_g.

The second algorithm tested for the minimization has been of a steepest descent type, in which the coordinate increments are given by

$$\delta \boldsymbol{X} = -\lambda \boldsymbol{G}_x \qquad (\lambda > 0),$$

with an expression for the step λ worked out using the second order variation of the energy. For instance, a partial contribution of second order variation (the complete expression should contain other terms, which have been omitted for the sake of simplicity) may be taken as

$$\delta E_s = \sum_{i=1}^{n_s} v_i \operatorname{tr}(\delta \boldsymbol{U} \boldsymbol{R}_i \delta \tilde{\boldsymbol{U}} \boldsymbol{f}) + \sum_{i \leqq j}^{n_s} v_{ij} \{\operatorname{tr}[\delta \boldsymbol{U} \boldsymbol{R}_j \delta \tilde{\boldsymbol{U}} \boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_i) + \delta \boldsymbol{U} \boldsymbol{R}_i \delta \tilde{\boldsymbol{U}} \boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_j)]\}.$$

Substituting $\delta \boldsymbol{U} = 4\lambda \boldsymbol{V}$, where $\boldsymbol{V} = \boldsymbol{P}^{-1} \boldsymbol{G}_x \boldsymbol{P}^{-1}$, in both the first and second order variations and equating to zero the derivative with respect to λ of the expression of the total variation, one gets for the step

$$\lambda_0 = \frac{m_1}{16 m_2},$$

where

$$m_1 = \operatorname{tr}(\boldsymbol{G}_x \tilde{\boldsymbol{G}}_x) \quad \text{and}$$

$$m_2 = \sum_{i=1}^{n_s} v_i \operatorname{tr}(\boldsymbol{V} \boldsymbol{R}_i \boldsymbol{V} \boldsymbol{f}) + \sum_{i \leqq j}^{n_s} v_{ij} \{\operatorname{tr}[\boldsymbol{V} \boldsymbol{R}_j \tilde{\boldsymbol{V}} \boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_i) + \boldsymbol{V} \boldsymbol{R}_i \tilde{\boldsymbol{V}} \boldsymbol{G}_{ij}(\bar{\boldsymbol{R}}_j)]\}.$$

The third algorithm of minimization tested makes a partial use of second derivatives. For small variations of $\boldsymbol{X}$ the condition of minimum can be put into the form

$$\left(\frac{\partial E}{\partial X_{kl}}\right)_0 + \sum_{r,s} \left(\frac{\partial^2 E}{\partial X_{rs} \partial X_{kl}}\right)_0 X_{rs} = 0 \qquad (k, l = 1, \dots m).$$

An approximate version of this expression is obtained making use only of the block diagonal part $\boldsymbol{D}_x$ of the second derivatives supermatrix, so that the coordinate increments may be approximated by

$$X_{kl} = -\left(\frac{\partial E}{\partial X_{kl}}\right)_0 \Bigg/ \left(\frac{\partial^2 E}{\partial X_{kl}^2}\right)_0 \qquad (k, l = 1, \dots m).$$

In processes of iteration, like those described, the inverse $\boldsymbol{I}$ of the matrix $\boldsymbol{P}$ need not be computed directly, but it may be obtained as the limit of the succession given by the recursion formula

$$\boldsymbol{I}_{n+1} = \boldsymbol{I}_n - 2\lambda_c \boldsymbol{P}(\boldsymbol{P} \boldsymbol{I}_n - \boldsymbol{1}). \tag{18}$$

As shown in Appendix 3, it is worked out using the steepest descent method. It has proved to be rapidly convergent, with the constant λ_c fixed equal to 0.5.

4. Preliminary Numerical Results

Some test calculations have been made on the molecules H_2S ($^1A_1$ state) and SO ($^3\Sigma^-$ state). For H_2S the following Slater type basis was used: $1s_H(1.2)$, $1s(15.5409)$, $2s(5.3144)$, $3s(2.1223)$, $2p(5.9885)$, $3p(1.8273)$, $3d(1.6)$, with geometry SH = 2.5427 a.u. and $H\hat{S}H = 119°48'$. For SO the basis on S was: $1s(15.5409)$, $2s(5.3144)$, $2p(5.9885)$, $3s(2.1223)$, $3p(1.8273)$, $3d(1.4)$, on O: $1s(7.6579)$, $2s(2.2458)$, $2p(2.30)$; the distance 3.40 a.u. These values (choosen only for the reason of having the corresponding integrals already computed) are taken from other calculations on minimization of expression (3) performed along the method of McWeeny (asymmetrical case) [1], with minor improvements [9]. All the calculations have been performed in single precision.

The algorithm tested more extensively is that of Murtagh and Sargent. The convergence rate is rather satisfactory, but, as expected, it becomes low near the minimum. Introducing a scale factor s_0 to multiply the coordinates X improves the convergence significantly. s_0 may be regarded as a kind of mean value of second derivatives, buth with such a constant scale factor at a certain iteration a divergence may arise because the descent step can turn into an ascent one: when this happens the initial value of s_0 (1.2, say) is decreased down to its normal value $s_0 = 1.0$. The convergence rate may also be improved by a proper choice of the parameter α, still keeping its value constant during a number of iterations. Table 1 illustrates these point for H_2S. The initial density matrix is obtained diagonalizing the core matrix and the energy associated is therefore the core energy. The values

Table 1. Minimization test for H_2S ($^1A_1$ state)

Iteration	α	s_0	Energy (a.u.)	m_g
0	0.0035	1.2	−396.080	86.3349
9	0.0035	1.0	−397.692	8.99128
62	0.0025	1.0	−397.784	5.66869
89	0.0085	1.0	−397.798	1.05417
108			−397.803	0.586050
0	0.0070	1.0	−397.743	14.3395
31			−397.814	0.369735

Table 2. Minimization test for SO ($^3\Sigma^-$ state)

Iteration	α	s_0	Energy (a.u.)	m_g
0	0.0035	1.1	−469.136	95.5178
13	0.030	1.0	−470.904	39.4416
32	0.0065	1.0	−471.112	1.29219
54			−471.133	0.578250

reported are only representative of an average behaviour: for instance, the rate of convergence between the 10th and 88th iterations may well be improved by changing the value of α, as it is shown in the second part of the table, where a different initial density matrix has been choosen. The same points are illustated in Table 2 for SO: here also the starting was taken deliberately far away diagonalizing the core matrix. In all the calculations the procedure has been found stable, even when working on the 7th significant figure.

To make a comparison, the rate of convergence of McWeeny procedure is by far lower and the minimum is attained at -397.787 a.u. for H_2S and at -471.125 a.u. for SO: numerical errors affect the 6th significant figure.

The other two algorithms tested have been the steepest descent and the second derivatives, along the formulation presented above. For both the rate of convergence has been found somewhat lower than in the preceding calculations. In particular, the rate obtained using the second derivatives according to the approximation described above does not seem to justify the necessary greater amount of computation, but this conclusion is by no means definitive because a systematic test on the use of second energy derivatives has still to be made, deserving a more careful investigation.

5. Conclusions

Our trial calculations have shown that the proposed method of incorporating constraints works rather satisfactory, but it seems worthwhile looking for minimization algorithms to speed up its rate of convergence particularly in the final stages. Perhaps it may be worth resorting to a systematic use of second derivatives.

But in this connexion some remarks are in order. It is to point out that within a minimization approach based on algorithms of the conjugate gradient type some difficulties of convergence are met [4], due to the possible singularity of the Hessian matrix, related to the presence of redundant variables. In the method presented in this paper the orthogonal transformation of the atomic basis requires just $m(m-1)/2$ variables and it seems likely that on the basis provided by these variables (i.e. the parameters $\boldsymbol{X}-\tilde{\boldsymbol{X}}$ of the orthogonal transformation) the Hessian should be free of singularities, because no redundant variable is present. Nevertheless, the energy is still invariant against a unitary transformation of the molecular orbitals of each shell, so that one is faced with the problem to remove this freedom reducing the number of the variables to the minimum just required [6] and to cast the method in terms of these variables.

Appendix 1

A. We wish to show that a density matrix can always be expressed in the form (8). Every $m \times m$ density matrix $\boldsymbol{R}$, due to its idempotency, may be written as $\boldsymbol{R}=\boldsymbol{T}\boldsymbol{T}^{\dagger}$, $\boldsymbol{T}$ being a $m \times m_0$ matrix (that relates the m_0 occupied molecular orbitals to the m atomic orbital basis) for which $\boldsymbol{T}^{\dagger}\boldsymbol{T}=\boldsymbol{1}_{m_0}$. Introducing the $m \times m$ unitary matrix $\boldsymbol{U}_0$ whose first m_0 columns are given by $\boldsymbol{T}$ and the $m \times m$ matrix $\boldsymbol{W}$ formed by the m_0 dimensional unit matrix $\boldsymbol{1}_{m_0}$ and null elements elsewhere, one has

$$\boldsymbol{R}=\boldsymbol{T}\boldsymbol{T}^{\dagger}=\boldsymbol{U}_0\boldsymbol{W}\boldsymbol{U}_0^{\dagger} \quad \text{and} \quad \boldsymbol{W}=\boldsymbol{U}_0^{\dagger}\boldsymbol{R}\boldsymbol{U}_0 .$$

Then, transforming by a new unitary matrix $\boldsymbol{U}_1$ completes the proof, namely

$$\bar{\boldsymbol{R}} = \boldsymbol{U}_1 \boldsymbol{W} \boldsymbol{U}_1^\dagger = \boldsymbol{U}_1 \boldsymbol{U}_0^\dagger \boldsymbol{R} \boldsymbol{U}_0 \boldsymbol{U}_1^\dagger = \boldsymbol{U} \boldsymbol{R} \boldsymbol{U}^\dagger ,$$

where $\boldsymbol{U} = \boldsymbol{U}_1 \boldsymbol{U}_0^\dagger$ is a unitary matrix[1].

B. In this section the equivalence of stationary conditions (15) with those of McWeeny method is shown. The minimum conditions obtained in [7] may be generalized as follows

(a) $\boldsymbol{R}' \boldsymbol{h}_i \boldsymbol{R}_i = \boldsymbol{0}$ $(i = 1, \dots n_s)$, where $\boldsymbol{R}' = 1 - \sum_{k=1}^{n_s} \boldsymbol{R}_k$.

(b) $\boldsymbol{R}_j (\boldsymbol{h}_i - \boldsymbol{h}_j) \boldsymbol{R}_i = \boldsymbol{0}$ $(i \neq j, i, j = 1, \dots n_s)$.

From these the conditions (15) may be obtained and reciprocally. From (a), summing over i and then subtracting from the equation so obtained its transpose one arrives at

$$\sum_{i=1}^{n_s} \boldsymbol{h}_i \boldsymbol{R}_i - \sum_{i=1}^{n_s} \boldsymbol{R}_i \boldsymbol{h}_i - \sum_{i,j(i \neq j)}^{n_s} \boldsymbol{R}_j (\boldsymbol{h}_i - \boldsymbol{h}_j) \boldsymbol{R}_i = \boldsymbol{0} .$$

Hence, taking into account (b) the (15) is obtained.

From the (15), multiplying by $\boldsymbol{R}'$ from the left (remembering that $\boldsymbol{R}' \boldsymbol{R}_i = \boldsymbol{0}, i = 1, \dots n_s$) and then by $\boldsymbol{R}_j$ from the right one gets (a) $\boldsymbol{R}' \boldsymbol{h}_j \boldsymbol{R}_j = \boldsymbol{0}$. Again, multiplying (15) by $\boldsymbol{R}_l$ from the left and then by $\boldsymbol{R}_k$ from the right one obtains (b) $\boldsymbol{R}_l (\boldsymbol{h}_k - \boldsymbol{h}_l) \boldsymbol{R}_k = \boldsymbol{0}$.

Appendix 2

The algorithm of the variable metric method of minimization proposed by Murtagh and Sargent [8] and its main features are summarized hereafter. The improved coordinates $\boldsymbol{X}_{n+1}$ are generated by the recursion formula

$$\boldsymbol{X}_{n+1} = \boldsymbol{X}_n - \alpha_n \boldsymbol{A}_n \boldsymbol{G}_{xn} ,$$

where α_n is a positive constant with values in the range $0 < \alpha_n \leqq 1$. The matrix $\boldsymbol{A}_n$ is initially set equal to the identity matrix ($\boldsymbol{A}_0 = \boldsymbol{1}$) and is updated by the formula

$$\boldsymbol{A}_n = \boldsymbol{A}_{n-1} + \boldsymbol{Z}_n \tilde{\boldsymbol{Z}}_n / c_n ,$$

where

$$\boldsymbol{Z}_n = \boldsymbol{A}_{n-1} [(1 - \alpha_{n-1}) \boldsymbol{G}_{xn-1} - \boldsymbol{G}_{xn}] , \qquad c_n = \operatorname{tr} [\widetilde{(\boldsymbol{G}_{xn} - \boldsymbol{G}_{xn-1})} \boldsymbol{Z}_n] .$$

To ensure stability the matrix $\boldsymbol{A}_n$ must be and remain positive definite: a simple test for this to occur is

$$b_n / c_n < 0 \quad \text{where} \quad b_n = \operatorname{tr} (\tilde{\boldsymbol{Z}}_n \boldsymbol{G}_{xn-1}) .$$

In case $\boldsymbol{A}_n$ becomes indefinite, it is set equal to the identity matrix. The method is quadratically convergent even if α_n is arbitrarily choosen and kept constant during all the process. The initial choice for α may be performed starting from, say, $\alpha_0 = 0.01$ and halfing down to a value which gives an energy and a gradient modulus less than the initial ones. An initial guess for α may be given by

$$0 < \alpha \leqq (E_n - E_{n-1}) / [\varepsilon \operatorname{tr} (\tilde{\boldsymbol{G}}_{xn} \boldsymbol{A}_n \boldsymbol{G}_{xn})] ,$$

where ε is a positive arbitrary constant < 1. It is shown in that if $b_n > 0$ for $\alpha_n = 1$, then the function has a minimum in the range $0 < \alpha_n < 1$. Other details are presented in the original paper.

Appendix 3

The inverse $\boldsymbol{I} = \boldsymbol{P}^{-1}$ of a non singular matrix $\boldsymbol{P}$ [in our case it is given by (2)] may be constructed iteratively by the steepest descent method. Let $\boldsymbol{I}_0$ a matrix which is only roughly equal to $\boldsymbol{P}^{-1}$. We seek a matrix $\delta \boldsymbol{I}_0$ such that $\boldsymbol{I}_0 + \delta \boldsymbol{I}_0$ is more nearly equal to $\boldsymbol{P}^{-1}$: the best possible choice of $\delta \boldsymbol{I}_0$ can be obtained by a steepest descent method, being only necessary to reduce to zero the quantity

$$\Omega = \operatorname{tr} [(\boldsymbol{P} \boldsymbol{I}_0 - 1) \widetilde{(\boldsymbol{P} \boldsymbol{I}_0 - \boldsymbol{1})}] .$$

---

[1] The author is indebted to Prof. E. Gianinetti for this proof.

The variation $\delta\Omega$ induced by a variation $\delta \boldsymbol{I}_0$ of $\boldsymbol{I}_0$ is given by a sum of a first and second order contributions. In first order

$$\delta\Omega_0 = \operatorname{tr}(\boldsymbol{B}\,\delta\tilde{\boldsymbol{I}}_0), \quad \text{where} \quad \boldsymbol{B} = 2\tilde{\boldsymbol{P}}(\boldsymbol{P}\boldsymbol{I}_0 - \boldsymbol{1}).$$

To reduce Ω most rapidly we therefore choose a step of length λ in the direction of steepest descent, that is

$$\delta\boldsymbol{I}_0 = -\lambda\boldsymbol{B} \qquad (\lambda > 0).$$

Taking now into account the second order contribution to $\delta\Omega$, λ is given by

$$\frac{\partial\delta\Omega}{\partial\lambda} = 0 = -\operatorname{tr}(\boldsymbol{B}\tilde{\boldsymbol{B}}) + 2\lambda\operatorname{tr}(\boldsymbol{P}\boldsymbol{B}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{P}}),$$

hence

$$\lambda_c = \frac{1}{2}\,\frac{\operatorname{tr}(\boldsymbol{B}\tilde{\boldsymbol{B}})}{\operatorname{tr}(\boldsymbol{P}\boldsymbol{B}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{P}})} \qquad (\lambda_c > 0).$$

From this one is lead to the simple iteration formula

$$\boldsymbol{I}_{n+1} = \boldsymbol{I}_n - 2\lambda_c\tilde{\boldsymbol{P}}(\boldsymbol{P}\boldsymbol{I}_n - \boldsymbol{1}),$$

which gives a sequence $\boldsymbol{I}_0, \boldsymbol{I}_1, \boldsymbol{I}_2 \ldots$ of increasingly good approximations to $\boldsymbol{P}^{-1}$. For the cases at hand, a constant value of $\lambda_c = 0.5$ has proved to give a very rapid convergence: in general no more than one or two iterations are needed.

Appendix 4

In this appendix the main steps for the derivation of a general expression of the second derivatives $\partial^2 E/\partial X_{mn}\partial X_{kl}$ are outlined. Their expression is worked out as a sum of all the contributions given by the coefficients of δX_{mn} in the matrix elements $(\delta \boldsymbol{G}_x)_{kl}$ obtained differentiating the kl element of the gradient matrix (14). These contributions arise from terms of two kinds, tipically of the form

$$(\boldsymbol{A}\,\delta\boldsymbol{X}\boldsymbol{B})_{kl} \quad \text{and} \quad [\chi_2(\mathscr{A}(\boldsymbol{1} \times \delta\boldsymbol{X}))]_{kl},$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are two matrices and $\mathscr{A}$ is a supermatrix (of dimensions $m^2 \times m^2$); the notation × for the direct (Kronecker) product has been introduced for brevity in order to obtain a final expression in a compact form. The simbol $\chi_2(\mathscr{C})$ represents the operation of contraction made on a supermatrix $\mathscr{C}$ and it is defined as the matrix of elements

$$[\chi_2(\mathscr{C})]_{rs} = \sum_t \mathscr{C}_{rt,st}.$$

It borns out because the matrix $\boldsymbol{G}_{ij}(\boldsymbol{R})$ may be written as

$$\boldsymbol{G}_{ij}(\boldsymbol{R}) = \chi_2[(\boldsymbol{1} \times \boldsymbol{R})\,g^{ij}],$$

g^{ij} being the supermatrix of elements given by the combination (5) of molecular two-electron integrals.

It is easy to see that the coefficients of δX_{mn} arising from the preceeding two differential forms are respectively given by

$$A_{km}B_{nl} = (\boldsymbol{A} \times \boldsymbol{B})_{kn,ml} \quad \text{and} \quad \mathscr{A}_{kn,lm}.$$

Defined the three supermatrices $\mathscr{G}$, $\mathscr{M}$, and $\mathscr{B}$ as follows:

$$\mathscr{G} = (\tilde{\boldsymbol{P}}^{-1} \times \boldsymbol{1})\,[(\boldsymbol{1} \times \tilde{\boldsymbol{P}}^{-1})(\boldsymbol{1} \times \boldsymbol{C}) - (\boldsymbol{C} \times \boldsymbol{1})(\boldsymbol{P}^{-1} \times \boldsymbol{1})]\,(\boldsymbol{1} \times \boldsymbol{P}^{-1}),$$

where

$$\boldsymbol{C} = \sum_{i=1}^{n_s} \boldsymbol{h}_i\bar{\boldsymbol{R}}_i - \sum_{i=1}^{n_s} \bar{\boldsymbol{R}}_i\boldsymbol{h}_i,$$

$$\mathscr{M} = (\tilde{\boldsymbol{P}}^{-1} \times \tilde{\boldsymbol{P}}^{-1})\left\{\sum_{i=1}^{n_s} [(\boldsymbol{1} \times \bar{\boldsymbol{R}}_i)\,\mathscr{H}_i - \mathscr{H}_i(\bar{\boldsymbol{R}}_i \times \boldsymbol{1})]\right\}(\boldsymbol{P}^{-1} \times \boldsymbol{P}^{-1})$$

where

$$\mathscr{H}_i = (\boldsymbol{h}_i \times \boldsymbol{1}) + (\boldsymbol{1} \times \boldsymbol{h}_i),$$

and

$$\mathscr{B}=(\tilde{\boldsymbol{P}}^{-1}\times\tilde{\boldsymbol{P}}^{-1})\left\{\sum_{i,j}^{n_s} v_i v_j[(\bar{\boldsymbol{R}}_i\times\mathit{1})f^{ij}-f^{ij}(\bar{\boldsymbol{R}}_i\times\mathit{1})]\right\}(\boldsymbol{P}^{-1}\times\boldsymbol{P}^{-1})$$

where

$$f^{ij}=(\mathit{1}\times\bar{\boldsymbol{R}}_j)\,g^{ij}-g^{ij}(\mathit{1}\times\bar{\boldsymbol{R}}_j)\,,$$

the final expression may be put into the form

$$\frac{\partial^2 E}{\partial X_{mn}\partial X_{kl}}=4(\mathscr{G}_{kn,ml}-\mathscr{G}_{km,nl})+8(\mathscr{M}_{km,nl}-\mathscr{M}_{kn,ml}+\mathscr{B}_{km,ln}-\mathscr{B}_{kn,lm})\,.$$

From the differential of (14) it is easy to see that the elements of $\mathscr{G}$ come from the differential terms associated with $\delta\boldsymbol{P}^{-1}$ and $\delta\tilde{\boldsymbol{P}}^{-1}$, those of $\mathscr{M}$ from $(\delta\bar{\boldsymbol{R}}_i)\,\boldsymbol{h}_i$ (and their transpose) and those of $\mathscr{B}$ from $\bar{\boldsymbol{R}}_i\delta\boldsymbol{h}_i$ (and their transpose).

Acknowledgements. Thanks are due to Professor M. Simonetta for his interest in this work. Thanks are due also to Dr. R. M. Stevens for his program on molecular integral calculation, and to Dr. M. P. Stabilini, who has adapted this program to a UNIVAC 1106 computer and has also contributed some sections of the SCF program. The author wishes to thank also Prof. E. Gianinetti for helpful discussions on several points.

References

1. McWeeny, R.: Proc. Roy. Soc. A **241**, 239 (1957)
2. Fletcher, R.: Mol. Phys. **19**, 55 (1970)
3. Kari, R., Sutcliffe, B.T.: Intern. J. Quantum Chem. **7**, 459 (1973)
4. Sutcliffe, B.T.: Theoret. Chim. Acta (Berl.) **33**, 201 (1974)
5. Raffenetti, R.C., Ruedenberg, K.: Intern. J. Quantum Chem. IIIs, 625 (1970)
6. Gianinetti, E.: Intern. J. Quantum Chem. **7**, 869 (1973)
7. McWeeny, R.: In: Molecular orbitals in chemistry, physics and biology. A tribute to R. S. Mulliken. New York: Academic Press 1964
8. Murtagh, B.A., Sargent, R.W.H.: Comp. J. **13**, 185 (1970)
9. The program illustration and the results will be published elsewhere

Prof. S. Polezzo
Istituto di Chimica Fisica dell' Università
Via Golgi 19
I-20133 Milano, Italy