

Optimization of Equilibrium Geometries and Transition Structures

H. Bernhard Schlegel*

Department of Chemistry, Wayne State University, Detroit, Michigan 48202
Received 13 July 1981; accepted 15 September 1981.

A modified conjugate gradient algorithm for geometry optimization is outlined for use with *ab initio* MO methods. Since the computation time for analytical energy gradients is approximately the same as for the energy, the optimization algorithm evaluates and utilizes the gradients each time the energy is computed. The second derivative matrix, rather than its inverse, is updated employing the gradients. At each step, a one-dimensional minimization using a quartic polynomial is carried out, followed by an *n*-dimensional search using the second derivative matrix. By suitably controlling the number of negative eigenvalues of the second derivative matrix, the algorithm can also be used to locate transition structures. Representative timing data for optimizations of equilibrium geometries and transition structures are reported for *ab initio* SCF-MO calculations.

INTRODUCTION

During the past decade, the analytical calculation of the first derivatives of the energy has become practical and efficient for *ab initio* molecular orbital methods.¹ For SCF calculations, analytical gradients can provide an order of magnitude more information about a molecule than the energy computation alone, while only doubling the computational time. Analytical derivative calculations recently have been extended to methods that include correlation energy.² Energy gradients were first used to calculate vibrational force constants.³ However, geometry optimization has now become the most widespread application. In this article we report a modified conjugate gradient algorithm for geometry optimization that is incorporated in the *ab initio* molecular orbital program GAUSSIAN 80.⁴ The algorithm takes into account the special economics of *ab initio* calculations, and is capable of converging on saddle points as well as minima.

From standard numerical analysis methods,⁵ a variety of algorithms for functional minimization are available that make effective use of gradients, e.g., conjugate gradients, variable metric, super memory gradient, and so forth. These techniques are all closely related and a number of applications to quantum mechanical optimization problems already exist.⁶ However, the established methods

can be improved by adapting them specifically to the unique features of *ab initio* calculations.

Energy calculations by *ab initio* methods require large quantities of computer time and central memory. Therefore, the optimization method can be quite elaborate and yet not significantly increase the overall demand on computer resources. Furthermore, the time to evaluate the *n*-dimensional gradient is frequently less than the energy computation time. Many standard algorithms assume the gradients require *n* times the effort of the function evaluation, based on computing the derivatives numerically and hence limit their use of the gradients. For *ab initio* optimizations, it is more efficient to calculate and use the gradient each time the energy is computed.

Most conjugate gradient algorithms work with the inverse of the hessian or second derivative matrix, and use a relatively simple updating scheme to improve the initial matrix. Usually this procedure must be restarted after *n* steps by reinitializing the hessian matrix. In the current application, one cannot afford to discard so much information. Additionally for geometry optimization, a great deal of information about the second derivative matrix is available from empirical rules and chemical principles. Thus, it is preferable to devise an algorithm based on the hessian, rather than its inverse, so that restarts can be avoided and so that chemical information can be incorporated readily.

* Alfred P. Sloan Fellow, 1981–1983.

ALGORITHM

Consider a function E , depending on n variables x_i , and the n -dimensional gradient of this function $g_i = dE/dx_i$. We wish to construct a series of steps leading to a stationary point of E , i.e., a point where $\mathbf{g} = 0$. Assume that p such steps have already been taken.

As step $p + 1$ we have the value of the function, E^a , and its gradient, \mathbf{g}^a , at $m + 1$ points ($0 \leq a \leq m \leq p$). Let this information be arranged such that \mathbf{x}^0 is the current point, \mathbf{x}^1 the most recent previous point, and so on, with \mathbf{x}^m the oldest. From the previous step we also have an $n \times n$ matrix \mathbf{F} that is an approximation to the true second derivative matrix, i.e., $F_{ij} = d^2E/dx_i dx_j$.

A single optimization step can be divided into three parts: (a) obtaining corrections to the approximate second derivative matrix, (b) searching for a minimum along the line between the current point and the previous point, and (c) estimating the location of the stationary point in the entire n -dimensional space by using the gradient and the approximate second derivative matrix. If no previous points are available, for example at step 1, parts (a) and (b) are omitted. At the beginning of the optimization process a suitable estimate of the second derivative matrix is required (*vide infra*).

Part a. Construct an orthonormal basis set of vectors \mathbf{r}^a for the space spanned by the vectors ($\mathbf{x}^a - \mathbf{x}^0$) by Schmidt orthogonalization.

$$\begin{aligned} \tilde{r}_i^a &= (x_i^a - x_i^0) - \sum_{b=1}^{a-1} r_i^b \sum_j (x_j^a - x_j^0) r_j^b \quad (1) \\ r_i^a &= \tilde{r}_i^a / \sqrt{\tilde{r}_i^a} \end{aligned}$$

Under certain circumstances, it may be necessary to eliminate some of the previous points. If a point is too far from the current point ($|\mathbf{x}^a - \mathbf{x}^0| > r_{\max}$) or if points are nearly collinear ($|\tilde{r}^a| < r_{\text{thresh}}$), then the point is discarded. Thus at step $p + 1$, a total of $m + 1$ energy and gradient evaluations remain for use in correcting the approximate force constant matrix.

In the space spanned by the vectors \mathbf{r}^a , obtain the new and the old estimates of the second derivatives, k_n^{ab} and k_o^{ab} , respectively:

$$k_n^{ab} = \frac{\sum_i (g_i^a - g_i^0) r_i^b - \sum_{c=1}^{a-1} k_n^{cb} \sum_i (x_i^a - x_i^0) r_i^c}{\sum_l (x_l^a - x_l^0) r_l^a} \quad (2)$$

$$k_o^{ab} = k_o^{ba} = \sum_{ij} r_i^a F_{ij} r_j^b \quad (3)$$

These values are then used to update the current estimate of \mathbf{F} .

$$F_{ij}^{\text{new}} = F_{ij}^{\text{old}} + \sum_{b \geq a}^m (k_n^{ab} - k_o^{ab}) [r_i^a r_j^b + (1 - \delta_{ab}) r_j^a r_i^b] \quad (4)$$

In eq. (2) the values for k_n^{ab} and k_n^{ba} will, in general, be different if E is not a quadratic function. However, we wish \mathbf{F} to be a symmetric matrix. The magnitude of the nonquadratic contributions to $(\mathbf{g}^a - \mathbf{g}^0)$ can be assumed to increase with the distance $|\mathbf{x}^a - \mathbf{x}^0|$. If we make the further assumption that each step predicts a point closer to the minimum, or reorder \mathbf{x}^a so that this is true, then $(\mathbf{g}^a - \mathbf{g}^0)$ are ordered approximately in terms of increasing nonquadratic contributions. Hence, for $b > a$, k_n^{ab} can be expected to have a smaller nonquadratic component than k_n^{ba} .

If the magnitude of the gradient vector is very large, it may be better to avoid updating the second derivative matrix until the gradients fall below a given cutoff. Thus the later steps in the optimization will not be spent correcting the bad approximation to the second derivatives obtained in the first few steps. The net effect of this approach is to let the first few steps be chosen in a steepest descent sense, until the gradients are reduced to a reasonable value.

Part b. Search for a minimum in E in the direction $(\mathbf{x}^a - \mathbf{x}^0)$ by fitting a polynomial to E^0 , \mathbf{g}^0 , E^1 , and \mathbf{g}^1 . In the current application, a quartic function is fitted to the two energies and directional gradients with the added constraint that $d^2E/dx^2 \geq 0$ with the equality holding at only one point. Such a curve has the advantage of possessing a single local minimum that is also the global minimum. This is found to be more satisfactory than a cubic polynomial. Let the coordinates of the minimum of the one-dimensional search be \tilde{x}_i . The derivatives \tilde{g}_i at \tilde{x}_i can then be obtained from \mathbf{g}^0 and \mathbf{g}^1 by interpolation.

Part c. Using the new estimate of the second derivative matrix \mathbf{F} from part a, and the result from the one-dimensional search of part b, the next estimate of the location of the stationary point is given by

$$x_i^{\text{new}} = \tilde{x}_i - \sum_j F_{ij}^{-1} \tilde{g}_j \quad (4)$$

The inverse of the second derivative matrix is calculated by first diagonalizing the matrix. The

Cols.	Description
123456789 ... etc.	
# HF/3-21G OPT	Procedure
BH4- H2O CLUSTER	Title
-1 1	Charge & Multip.
B	
O 1 BO	
H 1 BH1 2 OBH1	Z-Matrix
H 1 BH1 2 OBH1 3 180.	
H 1 BH2 2 OBH2 3 F90	
H 1 BH2 2 OBH2 3 -F90	
H 2 OH 1 BOH 3 0.	
H 2 OH 1 BOH 4 0.	
BO=3.2	Variables
BH1=1.25	
BH2=1.25	
OH=0.96	
OBH1=55.	
OBH2=125.	
BOH=54.	Constant
F90=90.	

Figure 1. Sample input for an optimization run. The geometry is specified via the *Z* matrix in terms of internal coordinates. For example, the sixth card of the *Z* matrix defines atom 6 as a hydrogen attached to atom 1, at a distance of BH2, making an angle with 2 of OBH2, and a dihedral angle 6-1-2-3 of -F90. Variables that are to be optimized are defined in a list following the *Z* matrix. Constants, if any, are specified in an additional list.

eigenvalues can then be tested to ensure that the matrix is positive definite and thus that the optimization step proceeds toward a minimum. If a negative eigenvalue is found, its sign is reversed. The net effect is to force a steepest descent step along the direction of the eigenvector possessing the negative eigenvalue. Additional bounds on the maximum and minimum magnitudes of the eigenvalues may be useful to improve the behavior of the optimization. The inverse matrix can then be constructed from the eigenvectors and eigenvalues. The total predicted change in any individual x_i resulting from both the linear and the quadratic search is required to be within certain limits. If the change in \mathbf{x} is too large, then the entire predicted change is scaled so that the limits are not exceeded.

The next step is carried out in a similar manner, retaining information from at most n previous steps and discarding the oldest information if necessary. At each step the gradient and displacement vectors are tested to see if the optimization has converged. The root-mean-square gradient and the absolute value of the largest component of the gradient must fall below their respective thresholds. The estimated displacement must pass a similar test on the root mean square and the absolute value of the largest component. If these four conditions are satisfied, the optimization is considered complete. If the conditions are

not met, the process is continued unless the maximum number of steps is exceeded.

After the implementation of the above approach, an optimization method by Brodli⁷ came to our attention. This algorithm also maintains the hessian instead of its inverse but uses a somewhat different updating scheme. For a quadratic function, Brodli's approach and most earlier algorithms require accurate one-dimensional searches in order to converge to the exact hessian. The present updating formula yields the exact hessian within the space spanned by the displacement vectors at any stage during the optimization, regardless of the manner in which the previous points are generated (albeit, at the cost of storing all the previous data). A further similarity is that both algorithms use a polynomial model for the one-dimensional search. Brodli has adopted a cubic polynomial, whereas we find a constrained quartic to be more suitable for minimizations. The fact that both methods perform well reinforces the merit and utility of the features common to the two algorithms.

IMPLEMENTATION

The algorithm outlined above is currently used in GAUSSIAN 80 to optimized molecular structures. This program,⁴ like its predecessors,⁸ employs the *Z*-matrix notation to input the geometry of a molecule using internal coordinates. The position of an atom is expressed in terms of bondlengths, valence angles, and dihedral angles relative to atoms already defined. The original *Z*-matrix definition⁸ has been extended to provide a simple, yet flexible method for specifying geometry optimizations, as illustrated in Figure 1. The internal coordinates in the *Z* matrix that are to be optimized are treated as algebraic variables and are given names which can be prefixed with a plus or minus sign. The initial value of each variable is specified in a parameter list appended to the *Z* matrix. The optimization of all coordinates or only a few parameters, the imposition of constraints on individual variables, and the utilization of symmetry are all readily achieved with this notation.

In addition to starting values of geometrical parameters, the optimization requires an initial estimate of the second derivative matrix. Various methods of increasing accuracy and expense are available within the framework of the GAUSSIAN 80 geometry optimization package. Bond stretching force constants can be estimated fairly

well from empirical rules,⁹ and average angle-bending and torsion constants can be obtained from vibrational spectra or theoretical calculations. Such an empirical guess at a diagonal force constant matrix is the default in GAUSSIAN 80. A more elaborate empirical force field similar to classical strain minimization calculations is a better starting point for optimizations of strongly coupled coordinates, such as those occurring in cyclic molecules.¹⁰ More accurate still is the direct calculation of a few second derivatives by finite differences. For each of the chosen coordinates, the molecule is displaced from its starting geometry and an additional calculation is performed. The diagonal second derivative can be obtained from an energy calculation at the displaced geometry combined with the energy and gradient at the starting point. If the gradient as well as the energy is calculated at the displaced geometry, then all the off-diagonal second derivatives associated with that coordinate can also be obtained. The best initial guess is a full force constant matrix, either read in from a previous calculation or calculated directly by numerical or analytical means.^{2a} Often it is useful to calculate the force constants with a smaller and cheaper basis set.

Typical examples of geometry optimizations* are collected in Table I. With reasonable guesses at the starting geometry, the number of optimization steps is usually fewer than the number of variables. Strongly coupled coordinates and very flexible modes of motion increase the number of steps.

TRANSITION STRUCTURES

A transition structure for a chemical reaction is a stationary point on an energy surface that is a local maximum along one and only one direction while being a local minimum in all other orthogonal directions. This unique direction is termed the transition vector and is the eigenvector associated with the negative eigenvalue of the force constant matrix. In general, the orientation of the transition vector is not known *a priori*, and hence must be determined in the course of a transition structure optimization. A recent review examines a variety of transition structure optimization methods¹²; most are unsatisfactory because of slow conver-

Table I. Sample geometry optimizations.^a

Molecule	Procedure	No. of Variables Optimized	No. of Steps	Timing Ratio Gradient Energy
BH ₄ ⁻ + H ₂ O ^c	HF/3-21G	7	7	0.63
BH ₄ ⁻ + H ₂ O ^d	HF/6-31G*	7	5	0.95
CH ₃ CFH	UHF/4-31G	15	11	0.70
para-xylene C ₈ H ₁₀	HF/STO-3G	11	8	1.61
CH ₃ PH ₃ ⁺	HF/4-31G	5	5	1.20
CH ₃ PH ₃ ⁺	HF/4-31G*	5	4	1.37
H ₂ C=CHSH	HF/4-31G	11	4	0.81 ^b
(CH ₃) ₂ S	HF/STO-3G	7	6	1.75 ^b
(CH ₃) ₂ SO	HF/4-31G + d on S	13	11	0.89 ^b
formyl guanidine (NH ₂) ₂ CNCHO	HF/STO-3G	19	17	1.67 ^b
formyl guanidine (NH ₂) ₂ CNCHO	HF/3-21G	19	10	0.38 ^b
CH ₃ OBH ₃ ⁻	HF/3-21G	13	8	0.67

^a Convergence thresholds were set so that the error in the final geometries are less than ± 0.001 Å for bondlength and $\pm 0.1^\circ$ for angles. Unless otherwise indicated, calculations were performed on the Amdahl 470V/6 at Wayne State University.

^b CDC Cyber 172 at Merck Sharp & Dohme using GAUSSIAN 78.

^c Starting with the geometry shown in Fig. 1.

^d Starting with the 3-21G optimized structure.

gence, lack of generality, or optimization to false stationary points.

The minimization algorithm outlined above can be modified to converge on transition structures as well, by forcing the second derivative matrix to contain a single negative eigenvalue. If certain initial conditions are met, the optimization approaches to the transition structure in the same manner as it proceeds to a minimum. The starting geometry must be chosen within the quadratic region of the transition structure, i.e., that region where the force constant matrix contains one negative eigenvalue. Additionally, the initial guess at the second derivative matrix must contain an appropriate negative eigenvalue. This is usually accomplished by calculating explicitly a few of the diagonal and off-diagonal force constants for the coordinates expected to dominate the transition vector, by using the finite-difference technique built into the minimization program. The updating scheme used to improve the second derivative matrix (part a above) automatically improves the direction of the transition vector, since it is an eigenvector of the matrix. At each step in the optimization, the eigenvalues of the force constant

* With this optimization method and an automatic archiving mechanism, it has been possible to optimize and tabulate more than 3000 molecules within the past two years. See ref. 11.

Table II. Sample transition structure optimization.^a

Molecule	Procedure	No. of Variables Optimized	No. of Steps		Timing Ratio Gradient Energy
			Prelim.	Optim.	
FCH ₂ CH ₂ --> FCHCH ₃	UHF/4-31G	15	3	11	0.63
FCH ₃ + ⁻ CCH F ⁻ +CH ₃ CCH	HF/4-31G	6	2	4	0.88
HOCH ₃ + ⁻ CN HO ⁻ +CH ₃ CN	HF/4-31G	11	2	4	0.77
CH ₂ +HF CH ₃ F	HF/3-21G	6	2	5	0.50

^a See footnote a of Table I. Preliminary steps were used to calculate certain elements of the force constant matrix. The total number of energy plus gradient computations is the sum of the preliminary and the optimization steps.

matrix are tested to ensure that a single negative eigenvalue is retained (see part c above). The one-dimensional search (part b) must be omitted, since it cannot be determined with any degree of certainty whether a minimum or a maximum is desired in that one direction. The second derivative matrix can be computed directly, if it is necessary to verify that the stationary point found is indeed a proper saddle point.

The present method does have a drawback in that the starting geometry must be in the quadratic region of the saddle point. For many reactions this is not a problem, since this region is relatively large and a reasonable starting structure can be estimated easily. However, there are some reactions such as those forbidden or nearly forbidden by orbital symmetry arguments, that may require a preliminary, manual survey of the energy surface to locate the quadratic region of the saddle point. A technique devised by Müller may be useful for this.¹³

Examples of transition structure optimizations are presented in Table II. The convergence to the saddle point is similar to convergence for minimization. The somewhat slower rate might be expected since transition states are generally more flexible than equilibrium structures, and since starting geometries and initial force constants cannot be estimated as well. Performance is enhanced further if the coordinates can be chosen so that the transition vector is dominated by only one or two coordinates.

Note added in proof. After the submission of this article, two additional publications appeared on finding transition states. Bell, Gordon, and Fletcher [*Chem. Phys. Lett.*, **82**, 122 (1981)] have modified a quasi-Newton (BFGS) algorithm to locate saddle points. A partitioning of the opti-

mization space is defined by the user and a maximum is found in one subspace, a minimum in the other. The main drawback is that the division of the optimization space is fixed, whereas in the present algorithm, it is improved in the course of the optimization. Cerjan and Miller [*J. Chem. Phys.*, **75**, 2800 (1981)] have devised a hill-climbing procedure for finding transition structures that requires analytical second derivatives at each step. Compared to the present work, a similar number of steps may be required, but each second derivative calculation requires roughly n times the computational effort of a gradient evaluation.

The author would like to thank Professor J. A. Pople and Professor L. C. Allen for constant encouragement during the development of these methods. He would also like to thank Professor Pople for an improvement to the algorithm. Financial support from the National Research Council of Canada and the Petroleum Research Foundation are gratefully acknowledged.

References

1. P. Pulay, in *Applications of Electronic Structure Theory*, H. F. Schaefer, III, Ed., Plenum, New York, 1977.
2. J. A. Pople, R., Krishnan, H. B. Schlegel, and J. S. Binkley, *Int. J. Quantum Chem.*, **13S**, 225 (1979); R. Krishnan, H. B. Schlegel, and J. A. Pople, *J. Chem. Phys.*, **72**, 4244 (1980); B. R. Brooks, W. D. Laidig, P. Saxe, J. D. Goddard, Y. Yamaguchi, and H. F. Schaefer, III, *J. Chem. Phys.*, **72**, 4652 (1980); S. Kato and K. Morkuma, *Chem. Phys. Lett.*, **65**, 19 (1979).
3. P. Pulay, *Mol. Phys.*, **17**, 197 (1969); D. M. Bishop and M. Randic, *J. Chem. Phys.*, **44**, 2480 (1966); S. Bratoz, *Coll. Int. C.N.R.S.*, **82**, 287 (1958).
4. J. S. Binkley, R. A. Whiteside, R. Krishnan, R. Seeger, D. J. DeFrees, H. B. Schlegel, S. Topiol, L. R. Kahn, and J. A. Pople, *GAUSSIAN 80, An Ab Initio Molecular Orbital Program*, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1980.
5. R. Fletcher, in *Numerical Methods for Unconstrained Optimization*, W. Murray, Ed., Academic, London, 1972; H. Y. Huang, *J. Opt. Theory Appl.*, **5**, 405 (1970), and references cited.
6. B. A. Murtagh and R. W. H. Sargent, *Comput. J.*, **13**, 185 (1970); J. W. McIver and A. Kormornicki, *Chem. Phys. Lett.*, **10**, 303 (1971); M. J. S. Dewar and Y. Yamaguchi, *Comput. Chem.*, **2**, 25 (1978); R. E. Kari, P. Mezey, and I. G. Csizmadia, *J. Chem. Phys.*, **63**, 581 (1975); D. Garton and B. T. Sutcliffe, in *Specialist Periodical Report, Theoretical Chemistry*, N. Dixon, Sr. Reporter, Chemical Society, London, 1974.
7. K. W. Brodli, *Math. Prog.*, **12**, 344 (1977), and references cited.
8. W. J. Hehre, W. A. Lathan, R. Ditchfield, M. D. Newton, and J. A. Pople, *GAUSSIAN 70*, Program No. 236, Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN, 1973.
9. R. M. Badger, *J. Chem. Phys.*, **2**, 128 (1934); **3**, 227 (1935).
10. H. B. Schlegel, unpublished.
11. R. A. Whiteside, J. S. Binkley, R., Krishnan, D. J. DeFrees, H. B. Schlegel, and J. A. Pople, *Carnegie-Mellon Quantum Chemistry Archive*, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1980.
12. K. Müller, *Angew. Chem. (Int. Ed. Engl.)*, **19**, 1 (1980), and references cited.
13. K. Müller and L. D. Brown, *Theor. Chim. Acta*, **53**, 75 (1979).