LOCATION OF STATIONARY POINTS OF POTENTIAL ENERGY SURFACES I. GUIDE TO A NEW PROGRAM SYSTEM DEVELOPED FOR THE DETERMINATION OF GRADIENT EXTREMAL CURVES

MIKLÓS. I. BÁN

Institute of Physical Chemistry, Attila József University

P.O.Box 105, H–6701 Szeged, Hungary

(Received October 10, 1989)

THE AIM OF THIS SERIES OF PAPERS IS TO PROVIDE THE INDISPENSABLE THEORETICAL BACKGROUND AND ALL PRACTICAL INSTRUCTIONS FOR USING A NEWLY DEVELOPED PROGRAM SYSTEM SUITABLE FOR THE EFFICIENT LOCATION OF SADDLE POINTS, MINIMA/MAXIMA AND POINTS OF GRADIENT EXTREMAL CURVES.

1. Introduction

Program system has been developed for the efficient location of saddle points, minima/maxima and points of gradient extremal curves[1]. It consists of two conceptually largely independent parts. Part I is designed for the determination of any kinds of stationary points of a function describing the potential energy surface of a chemical reaction. Part II relates to the determination of saddle points of gradient extremal curves.

The necessary theoretical considerations and technical instructions for the use of Part I are discussed in this paper, while Paper II will be devoted to the questions relating to Part II (see p.109 this volume).

Since a detailed theoretical discussion of the method has also been presented [1], the basic considerations concerning the algorithm will be mentioned here only when unavoidable. The FORTRAN code of the program is available; it will be submitted for distribution to the QCPE, but on request the author will send copies.

[1] The substantial comments provided in the program will also help the user.

2. The framework of the algorithm

To analyze the behaviour of the present algorithm, a quadratic objective function will be considered.:

$$f(x) = f(x_o) + g_o^t p + 1/2\, p^t G p \tag{1}$$

where x and x_o are the position vectors, $g_o = g(x_o)$ is the gradient, G is the matrix of second derivatives and $p = dx$ is the coordinate differential[2].

If the matrix of second derivatives (Hessian) is positive/negative definite, the stationary point of the function is a minimum/maximum. An indefinite Hessian determines a saddle surface and a saddle point for a critical point[3].

During the past two decades, the quasi–Newton algorithms[4] have proved to be very powerful in finding minima, and therefore they have also begun to be used for the location of saddle points [3]. If a genuine quasi–Newton method is applied to an indefinite objective function, several serious inherent problems occur.

With indefinite Hessians, singular vectors exist which satisfy the condition

$$v^t G v = 0. \tag{2}$$

Such singular directions of search cause the total breakdown of the algorithm. Neither the step size nor the update of the H–matrix is defined along a path of zero/very small curvatures. The existing techniques developed for handling such functions explicitly modify a parent quasi–Newton method for adaptation to indefinite optimization problems. The concept of the present algorithm approaches the problem of location of a saddle point in a different way. Instead of modifying a parent quasi–Newton algorithm, it seems more

[2]Small letters denote scalars and vectors, capital letters denote matrices, and superscript t refers to transposition.

[3] A saddle point can be of first or higher order, depending on the number of negative eigenvalues.

[4] For the common properties of the quasi–Newton family, see [2].

attractive to define a new (associated) surface, which is convex and contacts the original surface at its saddle point. This means that, in the sense of this idea, the two surfaces are in a first order touching contact characterized by identical function values and first derivatives, but their second derivatives differ in nature. While the critical point of the original surface is a saddle point, the "associated convex surface" has a minimum with the same coordinates. The introduction of the associated convex surface allows the use of any quasi–Newton method to find its minimum located at identical coordinates as the saddle point of the original surface. The present algorithm will carry this out.

The procedure is defined in an area around the saddle point where a constant direction vector (z_1) exists, along which the curvature of the surface is negative[5]:

$$z_1^t G z_1 < 0. \tag{3}$$

With the help of z_1 and $w_1 = Gz_1$ the reflector matrix

$$B = I - 2\frac{w_1 z_1^t}{z_1^t w_1} \tag{4}$$

can be constructed. The matrix B is used for the transformation

$$\tilde{g} = Bg, \tag{5}$$

the new gradient (5) being related to the convex surface[6].

If the algorithm is applied for the location of a minimum, the construction of (4)

[5] It has been presumed for chemical reasons that in the area to be explored the Hessian has only one negative eigenvalue.

[6] The transformation (5) of the gradient involves the implicit transformation

$$\tilde{G} = BG$$

of the Hessian and yields a positive definite quadratic form:

$$x^t \tilde{G} x > 0; \quad \forall x \in R^n.$$

and the transformation (5) are omitted. The further steps in both cases follow the general scheme of the quasi–Newton family. The minor changes introduced into the algorithm are of a practical nature and do not affect the common inherent structure of these procedures.

The peculiarities of the linear search, the properties of the update of the H–matrix and numerical examples will be presented in the forthcoming sections.

3. The linear search

The present description concerns the practical performance of the linear search step of a modified quasi–Newton algorithm suitable for the location of saddle points[7]. At any general point, the associated convex surface is defined exclusively on its derivative properties because the function value of the associated surface can not be evaluated from the primary data relating to the saddle surface[8].

The consequences of this fact are twofold:

i) the linear search must be based on the behaviour of the gradient, disregarding the continuous descent of the function value, which is an essential property of the quasi–Newton family, and

ii) the function value of the original (saddle–)surface becomes superfluous, saving part of the computational work.

In the search for a saddle point, the basic requirement is to know a constant vector (z_1) satisfying condition (3). The line segment connecting the minima at the "endpoints" of the reaction path is generally a suitable estimate.

At a point (x) of a general (non–quadratic) function the gradient and the (locally defined) B–matrix (4) have to be evaluated[9]:

---

[7]The essence of the necessary modifications relates to any member of the family, and therefore the method used will not be specified more closely.

[8]The function value is given only at the critical point.

[9]This is generally performed in a finite–difference approximation:

$$w_1(x) = g(x + \eta z_1) - g(x); \qquad \eta << 1.$$

$$B(x) = I - 2\frac{w_1(x)z_1^t}{z_1^t w_1(x)}. \tag{6}$$

The gradient will be transformed into the form

$$\tilde{g}(x) = B(x)g(x), \tag{7}$$

relating to the associated convex surface.

The direction of search (s) will be chosen according to the parent algorithm. The i^{th} search direction can generally be written as

$$s_i = -H_i\tilde{g}_{i,1}, \tag{8}$$

where H_i is a positive definite matrix, the double–index (i,1) is to be read "the gradient at point 1 along the i^{th} search direction" and the vector s_i must satisfy the condition

$$r_{i,1} = \tilde{g}_{i,1}^t s_i < 0. \tag{9}$$

The position (r) of a point is measured by the component of the gradient parallel to the path (directional derivative):

$$r \equiv r(x) = \tilde{g}^t s. \tag{10}$$

A point (x*) will be accepted as a local stationary point if the condition

$$r(x^*) = \tilde{g}^t(x^*)s = \epsilon \approx 0 \tag{11}$$

is fulfilled.

The step length (α) will be determined in a slightly different way if the point is on the side of descent/ascent around the local minimum, and therefore the linear search process will be discussed separately.

i) Side of descent:

The first step (at x_1) is

$$x_{i,2} = x_{i,1} + \alpha_{i,1} s_i \tag{12}$$

and the second step (at x_2) is

$$x_{i,3} = x_{i,2} + \alpha_{i,2} s_i. \tag{13}$$

The step length will be determined by using the information obtained on the curvature of the surface. The average curvature of the surface relating to the first step is

$$GS_{i,2} = (r_{i,2} - r_{i,1}) / \alpha_{i,1} \tag{14}$$

where $r_{i,2}$ and $r_{i,1}$ are the values of (9) at $x_{i,2}$ and $x_{i,1}$.

If

$$GS_{i,2} > 0, \tag{15}$$

the length ($\alpha_{i,2}$) of the step will be chosen as

$$\alpha_{i,2} = -r_{i,2} / (\omega_1 GS_{i,2}) \tag{16}$$

where

$$\omega_1 < 0. \tag{17}$$

The factor ω_1 accounts for the possible diminishing of the curvature and it is chosen for slight overestimation of the step length (generally, $\omega_1 = 0.8$).

If

$$GS_{i,2} < 0, \tag{18}$$

the step length will be estimated for a fixed factor ($\alpha_{i,2} = 4.0$).

The k^{th} step on the descent side at x_3 will be

$$x_{i,k+1} = x_{i,k} + \alpha_{i,k} s_i. \tag{19}$$

If both of the average curvatures ($GS_{i,k}$, $GS_{i,k-1}$) are positive, the change of the curvature will be approximated by

$$\omega_1^i = GS_{i,k}/GS_{i,k-1} \tag{20}$$

determining the step length:

$$\alpha_{i,k} = -r_{i,k}/(\omega_1^i GS_{i,k}). \tag{21}$$

If $GS_{i,k-1}$ is negative, $\alpha_{i,k}$ will be determined by (16) if $GS_{i,k}$ is positive, $\alpha_{i,k} = 4.0$.

This procedure is repeated until the actual point can be accepted as a local minimum on the descent side, or the point passes the minimum having reached the ascent side.

There are two criteria to be fulfilled at the local minimum. The first criterion requires an acceptable reduction of the gradient along the path. It is applied in two forms of optional use:

$$\text{a)}\ r_{i,k} < r_{i,1} \tag{22}$$

and

$$\text{b)}\ r_{i,k} < \omega_2 r_{i,1}. \tag{23}$$

The form condition (22) is a loose requirement used at the beginning of the iteration, while (23) is used if the point is closer to the minimum. The second criterion tends to ensure an acceptably close distance from the exact local minimum. This means that the step length of the k^{th} step will be reevaluated (and checked) through use of the data obtained in the k^{th} step:

$$GS'_{i,k} = (r_{i,k+1} - r_{i,k})/\alpha_{i,k}, \tag{24}$$

$$\alpha'_{i,k} = -r_{i,k}/GS'_{i,k}. \tag{25}$$

Before taking steps, the step length (21) is determined by using data on the curvature relating to the $(k-1)^{th}$ step (see(14)).

If

$$\alpha_k/\alpha'_k > \omega_3 \tag{26}$$

(generally, $\omega_3 = 0.8$), the point can be taken as acceptable.

If the first and second criteria are fulfilled simultaneously, the point will be accepted as a local stationary point. The last (approximating) point will always be stored.

ii) Side of ascent:

The $(k+1)^{th}$ point (reached in the k^{th} step and corresponding to x_4) is defined by (19). The local minimum has been passed:

$$r_{i,k} > 0, \quad (r_{i,l} < 0; \quad \forall l < k). \tag{27}$$

It will first be checked whether $x_{i,k}$ is too far from the minimum, *i.e.* the change of r must be smaller than a given limit:

$$|r_{i,k}|/|r_{i,k-1}| < \omega_4. \tag{28}$$

If this condition is not fulfilled, a step back (based on the measured curvature) will correct this situation. If the point is in a suitably close position, it will be stored (as will every successive approximation) and a bracketing process begins.

The $(l+2)^{th}$ point (corresponding to x_5 or x_5^l) is

$$x_{i,k+2} = x_{i,k} + \alpha_{i,k+1} s_i. \tag{29}$$

Using the (directional) first derivatives, the second derivatives will be approximated:

$$GG1 = (r_{i,k} - r_{i,k-1})/\alpha_{i,k-1}, \tag{30}$$

$$GG2 = (r_{i,k+1} - r_{i,k})/\alpha_{i,k}. \tag{31}$$

If both of them are positive, then an (approximate) third derivative is also computed:

$$GGG = (GG2 - GG1)/(\alpha_{i,k-1} + \alpha_{i,k}). \tag{32}$$

Because of the averaged nature of the second derivatives (GG1, GG2), the local value at point x_k will be approximated by the weighted sum:

$$GG(x_k) = GG_c = (\alpha^{-1}_{i,k-1} + \alpha^{-1}_{i,k})^{-1}(\alpha_{i,k-1} GG1 + \alpha^{-1}_{i,k} GG2). \qquad (33)$$

When the change in the function up to third order is taken into account, the two step lengths are:

$$(\alpha_{i,k+1}; \alpha'_{i,k+1}): \rightarrow (-GG_c \pm \sqrt{DISCR})GGG^{-1} \qquad (34)$$

where

$$DISCR = GG_c^2 - 2r_{i,k} GGG. \qquad (35)$$

The solution is chosen to be closer to

$$\alpha_{i,k+1} = -r_{i,k}/GG2 \qquad (36)$$

which provides the result of the second–order interpolation. If GG1 or DISCR is negative, (36) is used for estimation of the step length.

The new point ($x_{i,k+2}$ and also all subsequent points) can be on the descent side or the ascent side. After storage in the appropriate array, the bracketing process will be repeated, taking into account the points (k+2). If one of the two second derivatives GG2 is negative, the third–order interpolation will be substituted by the second–order step (36).

The conditions for exit are the same as discussed previously. The accuracy of the line search can be controlled via the parameter ω_2. The usual choice ($\omega_2 = 0.1$) prescribes a moderate accuracy, implying simultaneously a small number of gradient evaluations (generally in the range 1.8–2.3 per search direction).

4. Update of the H-matrix

To absorb the explored characteristics of the surface, the H–matrix will be updated

in each step. Because of the widely proved computational qualities, the present procedure allows the optional use of the DFP [4] and BFGS [5] update schemes.

The formulas

$$H_{i+1}^{DFP} = H_i^{DFP} + \frac{\delta_i \delta_i^t}{\tilde{\gamma}_i^t \delta_i} - \frac{(H_i \tilde{\gamma}_i)(\tilde{\gamma}_i^t H_i)}{\tilde{\gamma}_i^t H_i \tilde{\gamma}_i}, \tag{37}$$

$$H_{i+1}^{BFGS} = H_i^{BFGS} + \left[1 + \frac{\tilde{\gamma}_i^t H_i \tilde{\gamma}_i}{\tilde{\gamma}_i^t \delta_i}\right] \frac{\delta_i \delta_i^t}{\tilde{\gamma}_i^t \delta_i} - \frac{\delta_i (\tilde{\gamma}_i H_i) + (H_i \tilde{\gamma}_i) \delta_i^t}{\tilde{\gamma}_i^t \delta_i}. \tag{38}$$

These are formally strikingly similar to the original definitions in [4,5] ($\tilde{\gamma}_i$ and δ_i will be given below).

Since the coordinates and gradients relating to the last points (on both sides) are always preserved, there is some freedom in determining the curvature. Including the last point (accepted as a local minimum), there are altogether three points (and gradients) for this purpose (x_3, x_4, x_5, $(x_5^l)(k=3)$). Theoretical considerations and computational experience show that the best choice is the use of the weighted composition

$$\tilde{\gamma}_i = \tau_1(\tilde{g}_{i,k+2} - \tilde{g}_{i,k}) + \tau_2(\tilde{g}_{i,k+1} - \tilde{g}_{i,k+2}), \tag{39}$$

$$\delta_i = \gamma_1(x_{i,k+2} - x_{i,k}) + \tau_2(x_{i,k+1} - x_{i,k+2}) \tag{40}$$

where

$$\tau_1 = \alpha_{i,l+1}/\alpha_{i,l}, \tag{41}$$

$$\tau_2 = 1 - \tau_1. \tag{42}$$

The storage of the last points ensures a significant exclusion of higher than second–order effects.

The above update strategy yields a significant (30–40 percent) decrease in the number of steps (and gradient evaluations) necessary to attain the same accuracy as with the parent algorithm. The algorithm ensures a second–order convergence on quadratic surfaces.

5. The necessary conditions of the procedure

The explicit necessary condition is the knowledge of a vector (z_1) satisfying condition (3) in the whole region around the saddle point. Transformation (5) accounts for the reversion of the negative curvature of the surface (along z_1) into a positive one.

If the Hessian has more than one negative eigenvalue, the directions of negative curvature (conjugated to the Hessian) must be known to construct separate B–matrices.

Those areas where the Hessian has more than one negative eigenvalue are not of chemical importance, and therefore the present scheme of the algorithm covers the needs of chemical interest. The assumption that the Hessian has only one negative eigenvalue in the region of search is an implicit condition.

Acknowledgement

In writing up this paper thanks are due to Dr. Imre Bálint for the consent of using many results of his work, consultations and advice during the period working in my group between January 1985 and July 1987. The support of this work is acknowledged to the Hungarian Ministry of Higher Education (grant Nos. P–437/84. and 746/86).

References

[1] *Bálint I., M.I.Bán:* Theoret.Chim.Acta 63, 255 (1983),
Bálint, I.: Candidate dissertation, Szeged, 1986.

[2] *Fletcher, R:* Practical Methods of Optimization I.–II.,
Wiley, New York, 1980.
Powell, M.J.D.: Math.Progr.1,26 (1971).

[3] *Bálint,I.,M.I.Bán:* Int.J.Quant.Chem.25, 667 (1984).

[4] *Davidon,W.C.*: AEC Res.Dev.Rep.ANL–5990 (1959),
Fletcher,R.,M.J.D.Powell: Comp.J.6, 163 (1968).

[5] *Broyden,C.G.*: Maths. Comput.21, 368 (1967).
Fletcher,R.: Comp.J.13, 317 (1970).
Goldfarb,D.: Maths. Comput.24, 23 (1970).
Shanno,D.F.: Maths. Comput.24, 647 (1970).

РАСПОЛОЖЕНИЕ СТАЦИОНАРНЫХ ТОЧЕК ПОВЕРХНОСТЕЙ ПОТЕНЦИАЛЬНОЙ ЭНЕРГИИ I. ВВЕДЕНИЕ К НОВОЙ СИСТЕМЕ ПРОГРАММ ДЛЯ УСТАНОВЛЕНИЯ КРИВЫХ С ЭКСТРЕМАЛЬНЫМ ГРАДИЕНТОМ

М.И. БАН

Цель данной серии публикаций заключается в представлении необходимого теоретического обоснования и всех практических инструкций требуемых для пользования новь развитой системы программ пригодной для успешного установления седловных точек минимумов/максимумов и точек кривых с экстремальным градиентом.