Search for Stationary Points of Arbitrary Index by Augmented Hessian Method

YU. G. KHAIT

Russian Scientific Center "Applied Chemistry," Dobrolubova str. 14, 197198, St. Petersburg, Russia

A. I. PANIN

Department of Chemistry, Quantum Chemistry Group, St. Petersburg State University, University prospect 2, 198904, Old Petergof, St. Petersburg, Russia; E-mail: root@uni.chem.lgu.spb.su

A. S. AVERYANOV

Russian Scientific Center "Applied Chemistry," Dobrolubova str. 14, 197198, St. Petersburg, Russia

ABSTRACT _____

Convergence properties of the augmented Hessian (AH) method when searching for stationary points of an arbitrary fixed index are investigated. It is shown that the displacement vector of this method is proportional to one of the Hessian eigenvectors if the current point is far from a stationary one of the required index. A simple and reliable criterion for nearness of the current point to a stationary one of the desired index is proposed. The efficiency of a new one-dimensional optimization scheme that uses this criterion is studied. The case of coincidence of Hessian eigenvalues, which is a bottleneck of the standard AH method, is analyzed. A relation of the AH method to those by Poppinger and Wales is outlined. The correctness of the results obtained is illustrated on an example of a model surface. © 1995 John Wiley & Sons, Inc.

Introduction

During the last few years, notable attention has been paid to the development of methods for searching stationary points of different types (minima, maxima, saddle points) on the energy surface (see, e.g., [1-5]).

Among these methods, the augmented Hessian (AH) method proposed originally by Lengsfield [6] and developed later in [7-16] seems to be one of the most efficient and logically consistent. It is to be noted that it was the work by Jensen and Jorgensen [12] in which the AH method was developed for MCSCF optimization of excited states of molecular systems.

In this article, we consider those important features of the AH method that have not been displayed in the literature yet. We analyze, in particular, the relation of the AH method to those by Poppinger [17] and Wales [4, 18] and show that the displacement vector of the AH method is proportional to one of the Hessian eigenvectors if at the current point the Hessian is well conditioned and if this point is far from a stationary one of the required index. A simple and reliable criterion for nearness to a stationary point of given index is proposed. The efficiency of a one-dimensional optimization scheme that uses this criterion and that has been proposed by us earlier in [19] is investigated. The correctness of the results obtained is illustrated on an example of the model surface by Cerjan–Miller [20].

Basic Equations of the AH Method

We consider an energy surface defined by a twice continuously differentiable function $E(\mathbf{x})$ depending on n unconstrained variables $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. The problem is to find on such a surface a stationary point of index k, i.e., a point at which the second derivatives matrix (Hess matrix) of $E(\mathbf{x})$ is nondegenerate and possesses exactly k $(0 \leq k \leq n)$ negative eigenvalues.*

Let $\mathbf{x}_0 \in \mathbb{R}^n$ be an initial point and $h(\mathbf{x}_0)$ be the radius of the ball centered at this point,

$$\| \mathbf{x} - \mathbf{x}_0 \| \leq h(\mathbf{x}_0),$$

and it is supposed that within this ball the energy function $E(\mathbf{x})$ admits (with the required accuracy) the quadratic approximation:

$$E(\mathbf{x}) - E(\mathbf{x}_0) \simeq \Delta E^{(2)}(\mathbf{x}_0)$$

$$= \langle \mathbf{g} \mid \Delta \mathbf{x} \rangle + \frac{1}{2} \langle \Delta \mathbf{x} \mid \mathbf{H} \mid \Delta \mathbf{x} \rangle. \quad (1)$$

Here, $\Delta \mathbf{x} = \mathbf{x} - \mathbf{x}_0$ is a displacement vector and $\mathbf{g}$ and $\mathbf{H}$ are the gradient vector and the Hess matrix, respectively, calculated at point $\mathbf{x}_0$.

Introducing the augmented Hess matrix $\mathbb{H}$ of order $n + 1$,

$$\mathbb{H} = \begin{pmatrix} 0 & \mathbf{g}^+ \\ \mathbf{g} & \mathbf{H} \end{pmatrix},$$

*By definition, at a minimum point $k = 0$, at a saddle point (col) $k = 1, \ldots$, at a maximum point $k = n$.

we can rewrite expression (1) for $\Delta E^{(2)}$ in the form

$$\Delta E^{(2)}(\mathbf{x}_0) = \frac{1}{2} \langle 1, \Delta \mathbf{x} \mid \mathbb{H} \mid \begin{matrix} 1 \\ \Delta \mathbf{x} \end{matrix} \rangle$$

With such a representation for $\Delta E^{(2)}$, it seems natural to try to find an appropriate eigenvector of $\mathbb{H}$ for use it as a displacement vector. In fact, this is the primary idea of the AH method. We repeat briefly the main arguments leading to the basic equations of this method.

Let $\{\mathbf{u}_i, \varepsilon_i\}_{i=1}^n$ be orthonormal eigenvectors and eigenvalues of Hessian at point $\mathbf{x}_0$, and $\{\mathbf{w}_i, \lambda_i\}_{i=1}^{n+1}$ be the analogous set for the augmented Hessian $\mathbb{H}$. In accordance with the Cauchy "betweenness" theorem [21], the spectra of these matrices obey inequalities

$$\lambda_1 \leq \varepsilon_1 \leq \lambda_2 \leq \varepsilon_2 \leq \ldots \leq \varepsilon_k \leq \lambda_{k+1} \leq \varepsilon_{k+1} \leq \ldots$$
$$\leq \varepsilon_n \leq \lambda_{n+1}.$$

Let us analyze the displacement from $\mathbf{x}_0$ along the direction defined by $(k + 1)$st eigenvector of $\mathbb{H}$:

$$\mathbf{w}_{k+1} = \begin{vmatrix} a_0 \\ \mathbf{y}_{k+1} \end{vmatrix} \quad (0 \leq k \leq n), \quad (2)$$

where $\mathbf{y}_{k+1}$ is the projection of $\mathbf{w}_{k+1}$ on the coordinate space $\mathbb{R}^n$ (domain of the energy function), and a_0 is the projection of $\mathbf{w}_{k+1}$ on the orthogonal complement to this space in $\mathbb{R}^{n+1}$.

If a_0 differs from zero, the "augmented" displacement vector

$$\begin{vmatrix} 1 \\ \Delta \mathbf{x} \end{vmatrix} = \frac{1}{a_0} \mathbf{w}_{k+1} \quad (3)$$

from $\mathbf{x}_0$ may be used and, due to the normalization condition

$$a_0^2 + \|\mathbf{y}_{k+1}\|^2 = 1,$$

the following quadratic change of the energy corresponds to this displacement:

$$\Delta E^{(2)}(\mathbf{x}_0) = \frac{\lambda_{k+1}}{2a_0^2}. \quad (4)$$

Thus, under small displacements along the eigenvector $\mathbf{w}_{k+1}$, the sign of the energy shift is determined by the sign of the $(k + 1)$st eigenvalue of the AH.

Let us make out under what conditions a_0 differs from zero. Using representation (2), is it easy to rewrite the eigenvalue equation

$$\mathbb{H}\, \mathbf{w}_{k+1} = \lambda_{k+1} \mathbf{w}_{k+1},$$

in the form

$$\begin{cases} \langle \mathbf{g} \mid \mathbf{y}_{k+1} \rangle = a_0 \lambda_{k+1} & (5a) \\ (\mathbf{H} - \lambda_{k+1})\mathbf{y}_{k+1} + a_0 \mathbf{g} = 0 . & (5b) \end{cases}$$

From Eq. (5b), it immediately follows that if both $\mathbf{x}_0$ is not a stationary point ($\|\mathbf{g}\| \neq 0$) and strict inequalities

$$\varepsilon_k < \lambda_{k+1} < \varepsilon_{k+1} \qquad (6)$$

keep true, then $a_0 \neq 0$ (proof ad absurdum) and, as it is readily seen from Eqs. (3) and (5), the equalities

$$\lambda_{k+1} = -\langle \mathbf{g} \mid (\mathbf{H} - \lambda_{k+1})^{-1} \mid \mathbf{g} \rangle , \qquad (7)$$

$$\Delta \mathbf{x} = \frac{1}{a_0} \mathbf{y}_{k+1} = -(\mathbf{H} - \lambda_{k+1})^{-1} \mathbf{g} \qquad (8)$$

are valid. Indeed, if inequalities (6) are satisfied, then λ_{k+1} does not belong to the spectrum of $\mathbf{H}$, matrix $\mathbf{H} - \lambda_{k+1}$ is nondegenerate and possesses the required number k of negative eigenvalues, and, as a result, the displacement vector $\Delta \mathbf{x}$ can be readily determined from Eq. (8). This equation is also useful for qualitative characterization of displacements recommended in the AH method. Indeed, far from the solution, where the Hessian at current point $\mathbf{x}_0$ does not possess the required number k of negative eigenvalues, two cases are to be treated: (1) $\varepsilon_k < \lambda_{k+1} < \varepsilon_{k+1} < 0$, and (2) $0 < \varepsilon_k < \lambda_{k+1} < \varepsilon_{k+1}$. In the first case, $\mathbf{H}(\mathbf{x}_0)$ has a "superfluous" number of negative eigenvalues and $\mathbf{x}_0$ is situated certainly higher [in $E(\mathbf{x})$ scale] than the solution desired. Due to Eq. (4), displacement (8) corresponds to the descent down the energy surface to the region where $\mathbf{H}$ *may possess* less in number negative eigenvalues than that at the starting point $\mathbf{x}_0$. In the second case, the point $\mathbf{x}_0$ is lower than the desired stationary point, displacement vector (8) corresponds to the ascent up the energy surface, and one *may hope* that the number of negative eigenvalues at the new point (after displacement from $\mathbf{x}_0$) will be greater than that at the starting point.

If $\mathbf{x}_0$ is sufficiently close to the solution, then the Hess matrix $\mathbf{H}(\mathbf{x}_0)$ possesses the required number k of negative eigenvalues and the following strict inequalities should hold true:

$$\varepsilon_k < 0 < \varepsilon_{k+1} .$$

In this case, as seen from Eqs. (7) and (8), when $\|\mathbf{g}\| \to 0$, λ_{k+1} goes to zero as

$$|\lambda_{k+1}| \le \|\mathbf{g}\| \circ \|\Delta \mathbf{x}\| \le \|(\mathbf{H} - \lambda_{k+1})^{-1}\| \circ \|\mathbf{g}\|^2$$

$$\times \longrightarrow 0 ,$$

and Eq. (8) automatically reduces to the standard Newton–Raphson equation.

In concluding this section it is to be noted that all the results described here are well known primarily due to [9,12,13].

Choice of AH Displacement Vector

When a current point $\mathbf{x}_0$ is far from the solution, then, under certain additional conditions, the displacement used in the AH method may be characterized in more detail. These conditions are that (1) the Hessian eigenvalues ε_k and ε_{k+1} should not coincide ($\varepsilon_k < \varepsilon_{k+1}$) and (2) these eigenvalues should be sufficiently large (in comparison to the gradient components). Indeed, as seen from Eq. 5(b), under such conditions, the eigenvalue λ_{k+1} of $\mathbf{H}$ is close to either ε_k or ε_{k+1}, and the vector $\mathbf{y}_{k+1}$ is close to the corresponding eigenvector of Hessian ($\mathbf{u}_k$ or $\mathbf{u}_{k+1}$). Therefore, rewriting the right-hand side of Eq. (7) in the basis of the Hessian eigenvectors, we can omit all the arising terms except for one. Solving the thus-obtained simplified equation, we arrive at the following estimations for λ_{k+1} [that satisfy inequalities (6)]: In the case $\varepsilon_k < \varepsilon_{k+1} < 0$:

$$\lambda_{k+1} = \frac{1}{2} \varepsilon_{k+1}\left(1 + [1 + 4g_{k+1}^2/\varepsilon_{k+1}^2]^{1/2}\right); \qquad (9)$$

and in the case $0 < \varepsilon_k < \varepsilon_{k+1}$;

$$\lambda_{k+1} = \frac{1}{2} \varepsilon_k\left(1 + [1 + 4g_k^2/\varepsilon_k^2]^{1/2}\right), \qquad (10)$$

where $g_i = \langle \mathbf{g} \mid \mathbf{u}_i \rangle$ ($i = 1, 2, \ldots, n$).

Expressions (9) and (10) had been proposed earlier by Wales [18] for calculation of shift parameters on the base of certain physical arguments when the movement is performed along the eigenvectors of the (nonaugmented) Hessian. It is clear now that, in fact, Wales used implicitly an approximation for AH eigenvalues as shift parameters.

The concrete calculations show that these estimations are sufficiently reliable if $\mathbf{H}$ does not possess small eigenvalues. In such a case, it is possible to obtain also a good approximation for $(k + 1)$st eigenvector of $\mathbf{H}$. Indeed, let us recast system (5) as

$$\begin{cases} \mathbf{y}_{k+1} = \dfrac{1}{\lambda_{k+1}}\left(\mathbf{H}\mathbf{y}_{k+1} + \dfrac{\langle \mathbf{g} \mid \mathbf{y}_{k+1} \rangle}{\lambda_{k+1}} \mathbf{g}\right) \\ a_0 = \dfrac{\langle \mathbf{g} \mid \mathbf{y}_{k+1} \rangle}{\lambda_{k+1}} . \end{cases} \qquad (11)$$

As seen from Eqs. (9) and (10), augmentation of $\mathbf{H}$ results in [see Eq. (4)] negative shifts for negative

eigenvalues and positive shifts for positive ones. In the case $0 < \varepsilon_k < \varepsilon_{k+1}$, the eigenvalue is λ_{k+1} "genealogically" related to ε_k, and we can use $y_{k+1} = u_k$ as the initial approximation in handling Eq. (11). If ε_k is sufficiently large, then, with the aid of Eq. (10), system (11) may be reduced to

$$\begin{cases} y_{k+1} \simeq \left(u_k + \dfrac{g_k}{\varepsilon_k^2} g \right)\left(1 - \dfrac{g_k^2}{\varepsilon_k^2} \right) \\ a_0 \simeq \dfrac{g_k}{\varepsilon_k}\left(1 + \dfrac{\|g\|^2}{\varepsilon_k^2} \right). \end{cases}$$

In the case

$$\left| \frac{g_k}{\varepsilon_k} \right| \ll 1, \tag{12}$$

the last equations take the following simple form[†]:

$$a_0 = \frac{g_k}{\varepsilon_k},$$

$$y_{k+1} = u_k,$$

$$\lambda_{k+1} = \varepsilon_k(1 + a_0^2), \tag{13}$$

and expression (8) for the displacement vector becomes

$$\Delta x = \frac{\varepsilon_k}{g_k} u_k. \tag{14}$$

Several interesting conclusions can be made on the basis of these equations, namely, if the stationary point of index k is to be reached from a certain point x_0 at which the Hessian possesses "redundant" positive eigenvalues (more than $n - k$), then, in accordance with Eq. (14), the movement should be performed along the kth eigenvector u_k of the Hessian $H(x_0)$ and it corresponds to the ascent up the energy surface $(\lambda_{k+1} > 0)$. In particular, if we are in some vicinity of a minimum point $(k = 0)$ and are searching for a saddle point $(k = 1)$, then the movement is performed along the lowest eigenvector u_1 of the Hessian. For example, the initial displacement from the equilibrium molecular geometric structure to a transition state will be along the vibrational mode with the lowest frequency, i.e., along the weakest bond. This confirms the suggestion posed by Poppinger in 1975 [17].

In an analogous manner, the case $\varepsilon_k < \varepsilon_{k+1} < 0$ may be treated. In this case, the Hessian at a current

point has several "redundant" negative eigenvalues (more than k) and λ_{k+1} is "genealogically" related to ε_{k+1} [see Eq. (9)]. The corresponding equations may be written in the form

$$a_0 = \frac{g_{k+1}}{\varepsilon_{k+1}},$$

$$y_{k+1} = u_{k+1},$$

$$\lambda_{k+1} = \varepsilon_{k+1}(1 + a_0^2),$$

$$\Delta x = \frac{\varepsilon_{k+1}}{g_{k+1}} u_{k+1} = -\frac{|\varepsilon_{k+1}|}{g_{k+1}} u_{k+1}. \tag{15}$$

In this case, movement to a stationary point of index k is performed along $(k + 1)$st eigenvector u_{k+1} *without dependence on the full number of negative eigenvalues of the Hessian*. This movement corresponds to the descent down the energy surface $(\lambda_{k+1} < \varepsilon_{k+1} < 0)$. For example, if we are in some vicinity of a maximum point, then we move along either u_1 or u_2 when searching for a minimum $(k = 0)$ or a saddle $(k = 1)$ point, respectively. It is to be noted that, according to Poppinger [17], in this case, the movement should be performed along the Hessian eigenvector corresponding to the smallest (by absolute value) negative eigenvalue. We see that such a strategy can hardly be recommended as universal.

Now let us turn to a more complicated case: $\lambda_{k+1} = \varepsilon_k = \varepsilon_{k+1}$. [‡]In this case, the displacement vector cannot be determined from Eq. (8) since the matrix $H(x_0) - \lambda_{k+1}$ is not invertable. We should therefore return to the initial system (5) which, in the particular case under consideration, has the following exact solution:

$$a_0 = 0, \tag{16}$$

$$y_{k+1} = [g_k^2 + g_{k+1}^2]^{-1/2}(g_{k+1}u_k - g_k u_{k+1}). \tag{17}$$

We cannot use Eq. (3) for calculation of the displacement vector and should appeal to the original strategy of the AH method considering displacements along y_{k+1} with a certain step γ:

$$\Delta x = \gamma \circ y_{k+1}. \tag{18}$$

In accordance with Eq. (1), small displacements along y_{k+1} correspond with the following quadratic energy change:

$$\Delta E^{(2)}(x_0) = \frac{1}{2} \gamma^2 \lambda_{k+1}.$$

[†]These equations may be obtained by means of standard perturbation theory [22], if in the matrix equality

$$H = \begin{pmatrix} 0 & 0 \\ 0 & H \end{pmatrix} + \begin{pmatrix} 0 & g^+ \\ g & 0 \end{pmatrix},$$

the second term on its right-hand side is considered as perturbation. We will use this approach hereafter.

[‡]Since, by assumption, we consider only stationary points, where $\varepsilon_k < 0 < \varepsilon_{k+1}$, the degenerate case, $\varepsilon_k = \varepsilon_{k+1}$, may occur only far from the solution.

Note that the sign of γ in Eq. (18) cannot be determined using the quadratic approximation for energy since the displacement vector (18) is orthogonal to the gradient vector.

If the current point x_0 is close to the desired stationary point (in particular, $\varepsilon_k < 0 < \varepsilon_{k+1}$ in this point) and the Hessian is well conditioned, then, treating $\begin{pmatrix} 0 & g^+ \\ g & 0 \end{pmatrix}$ as perturbation, it is easy to get the estimation for $(k + 1)$st eigenvalue of $\mathbb{H}$:

$$\lambda_{k+1} = -\sum_{i=1}^{n} g_i^2/\varepsilon_i,$$

which is valid up to the third order of the perturbation theory [22].

For the components of the eigenvector w_{k+1}, we have the following first-order estimation:

$$a_0 = 1$$
$$y_{k+1} = -\sum_{i=1}^{n} (g_i/\varepsilon_i)u_i. \tag{19}$$

Therefore, when close to the solution, the strategy of search in the AH method is equivalent to that in the Newton–Raphson method:

$$\Delta x = -\sum_{i+1}^{n} (g_i/\varepsilon_i)u_i,$$
$$\Delta E^{(2)}(x_0) = -\frac{1}{2}\sum_{i=1}^{n} g_i^2/\varepsilon_i.$$

Using the results obtained, it is possible to formulate a criterion for a point x_0 to be close to the desired stationary point [if, of course, $H(x_0)$ is well-conditioned]. As a quantative measure of such nearness, the absolute value of $|a_0|$ is either small [see Eqs. (13) and (15)] or equal to zero [in the degenerate case, $\varepsilon_k = \varepsilon_{k+1}$, see Eq. (16)]. On the contrary, in a vicinity of a solution, $|a_0|$ is close to 1 [see Eq. (19)]. In the next section, we demonstrate how to exploit this "nearness criterion" on a one-dimensional optimization level.

One-Dimensional Optimization in the AH Method

It is well known that the convergence properties of the iterative optimization process for non-quadratic functions depend essentially on a strategy of step-length adjustment, when moving along the current direction of the search. At present,

the step-length adjustment algorithm proposed by Fletcher [23] is widely used. This algorithm eliminates the possibility of large steps along displacement vectors by imposing the additional restriction that the final displacement should be inside the ball of radius $h(x_0)$ within which the quadratic approximation of the object function is valid (with the required accuracy, of course).

Fletcher's approach in the frameworks of the AH method would generate the displacement vector

$$\Delta x = \gamma \circ y_{k+1}/\|y_{k+1}\|, \tag{20}$$

where $\gamma = h(x_0)$.

In the AH method, however, another approach proposed by Jensen et al. [12, 16] is exploited. It consists of replacing the AH matrix $\mathbb{H}$ by the parametric one:

$$\mathbb{H}^\alpha = \begin{pmatrix} 0 & \alpha g^+ \\ \alpha g & H \end{pmatrix},$$

where parameter α is to be determined from the equation

$$\|\Delta x^\alpha\| = h(x_0). \tag{21}$$

Here, the displacement vector Δx^α is connected with the components of the $\mathbb{H}^\alpha$ eigenvector

$$w_{k+1}^\alpha = \left| \begin{array}{c} a_0^\alpha \\ y_{k+1}^\alpha \end{array} \right\rangle$$

by the relation

$$\Delta x^\alpha = \frac{1}{\alpha a_0^\alpha} y_{k+1}^\alpha = -(H - \lambda_{k+1}^\alpha)^{-1}g.$$

The disadvantage of this approach is in the lack of explicit (analytic) dependence of y_{k+1}^α and a_0^α on parameter α, which makes the solution of Eq. (21) somewhat complicated. Let us discuss this problem in more detail.

Using the same arguments as in the previous section leads to the conclusion that if (i) x_0 is far from the solution, (ii) H is well conditioned, and (iii) $\varepsilon_k \neq \varepsilon_{k+1}$, then the equations

$$a_0^\alpha = \alpha \circ a_0,$$
$$y_{k+1}^\alpha = y_{k+1},$$
$$\Delta x^\alpha = \frac{1}{\alpha^2 a_0} y_{k+1}$$

are valid. On the right-hand sides of these equations, the components of the nonparametrized AH appear. Using the last expression for the displacement vector, it is easy to get the solution of Eq. (21). Such an approach, however, also possesses certain

drawbacks: First of all, it requires preliminary calculation of y_{k+1} and a_0. Then, it is not universal, because it is not difficult to show that in the cases when either $\varepsilon_k = \varepsilon_{k+1}$ and x_0 is far from the solution or x_0 is close to the solution the choice of α does not affect the step length. Therefore, in such cases, we should turn to the displacement vector (20) with $\gamma = h(x_0)$.

An alternative scheme for step-length adjustment in the AH method has been proposed in our earlier article [19]. The concrete optimization algorithm depends on how far the current point x_0 is from the solution. In the previous section, it was shown that as a measure for x_0 to be near the stationary point of index k the absolute value of the a_0 component of w_{k+1} may be used.

If $|a_0|$ is sufficiently large (in practice, "sufficiently large" means $|a_0| \geq 0.75$), then x_0 is close to the solution and the optimal step length γ_0 may be obtained from the condition[§]

$$|\langle g(\gamma_0)|y_{k+1}\rangle| = \min_{\gamma \geq 0} |\langle g(\gamma)|y_{k+1}\rangle|, \quad (22)$$

where $g(\gamma)$ is the gradient vector calculated at the point $x_0 + \Delta x(\gamma)$.

If $|a_0|$ is relatively small ($|a_0| < 0.75$), then the current point seems to be far from the solution. It may be expected that in some vicinity of x_0 the Hessian does not have the required number of negative eigenvalues and, therefore, it is desirable to leave this region as soon as possible. Since vector y_{k+1} generated in the AH method supplies us with quite a reasonable direction of movement and the gradient vector points in the direction of steepest energy change, it seems to be sound to move along y_{k+1} as far as the angle between y_{k+1} and $g(\gamma)$ decreases. This criterion for the step-length adjustment may be written in the form

$$\frac{|\langle g(\gamma_0)|y_{k+1}\rangle|}{\|g(\gamma_0)\|} = \max_{\gamma \geq 0} \frac{|\langle g(\gamma)|y_{k+1}\rangle|}{\|g(\gamma)\|} \quad (23)$$

Applications

To test the correctness of our conclusions and the efficiency of step-length adjustment algorithms, we chose the well-known model surface that had been investigated earlier in [13,20,24]:

$$f(x,y) = (a - by^2)x^2 \exp(-x^2) + 0.5cy^2. \quad (24)$$

[§]It is clear that condition (22) is equivalent to the criterion of minimum for the gradient norm square along the line y_{k+1}.

For positive parameters a, b, and c, this function has a maximal number of stationary points if

$$b > 0.5c \circ \exp(1).$$

For our analysis, we took values $a = 1$, $b = 1.5$, and $c = 1$, satisfying the last inequality. In such a case, $f(x,y)$ has four accessible stationary points in the first quadrant: minimum point $A = (0,0)$, two saddle points $B = (0.7868, 0.8165)$ and $C = (1.2297, 0.8165)$, and maximum point $D = (1,0)$. Figure 1 displays a map of $f(x,y)$ and its stationary points in the first quadrant.

We used three versions of the AH method to localize stationary points on the model surface (24). In all versions, Eq. (20) was employed for calculation of displacement vectors and step-length adjustment was performed in accordance with algorithms based on Eqs. (22) and (23). In the first two versions, the component y_{k+1} of the $(k + 1)$st eigenvector w_{k+1} of the AH H was used to generate the direction of movement. The difference between these two versions was in employing different nearness criteria between current and stationary points. In the first version, this criterion was based on Hessian spectrum analysis, while in the second case, the $|a_0|$ value was used in the manner described in the previous section. In the third version, again, the Hessian spectrum was exploited to control the nearness of x_0 to a stationary point, but the generation of vector y_{k+1} was performed in accordance with estimations obtained in the third section, namely, far from the solution, we took $y_{k+1} = g_k u_k$ if $0 < \varepsilon_k < \varepsilon_{k+1}$, $y_{k+1} = -g_{k+1}u_{k+1}$ if $\varepsilon_k < \varepsilon_{k+1} < 0$, and near the solution, the Newton–Raphson displacement vector was generated. If far from the solution the case $\varepsilon_k = \varepsilon_{k+1}$ occurred, then y_{k+1} was evaluated in accordance with the strict equality (17).

The results of test calculations for several starting points are presented in Table I. From this table, it is seen that all three versions of the AH method display satisfactory convergence properties. The actual character of the optimization process is in very good agreement with theoretical conclusions of the third section. Both our step-length adjustment algorithms and nearness criterion based on the $|a_0|$ value also proved to be sufficiently reliable and efficient.

At the same time, it is the second version of the AH method that seems to be the simplest from the practical viewpoint. In this version, (i) the component y_{k+1} of the AH eigenvector w_{k+1} is taken as a direction of movement, (ii) the displacement vector Δx_{k+1} is generated in accordance with Eq. (20), and
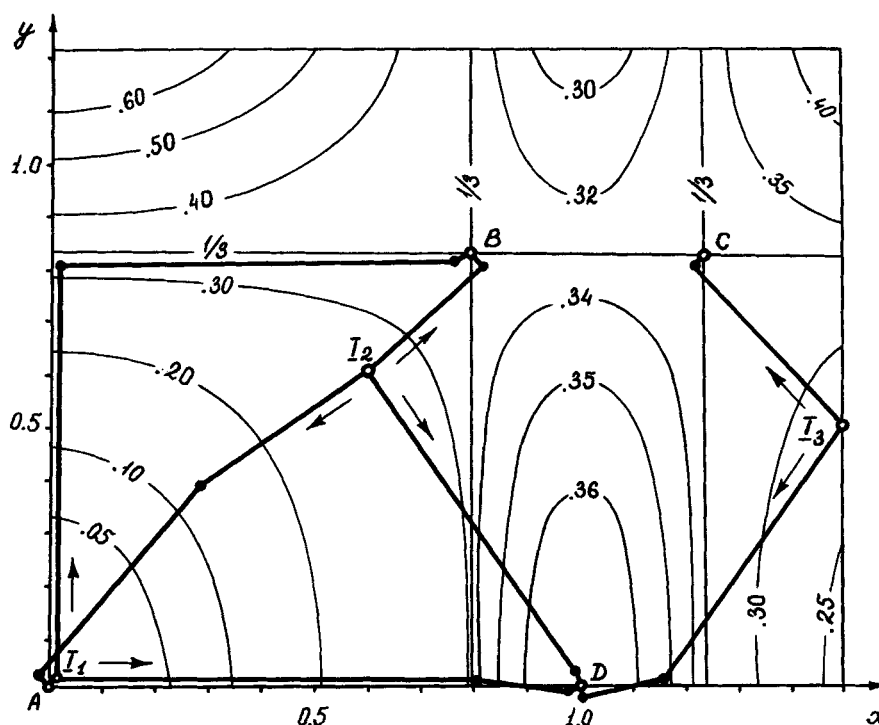
FIGURE 1. The convergence paths of the recommended (second) version of the AH method from starting points $I_1 = (0.01,0.01)$, $I_2 = (0.6,0.6)$, and $I_3 = (1.5,0.5)$ to different stationary points on the Cerjan–Miller surface.

(iii) step length γ is adjusted in accordance to either condition (22) or condition (23) depending on the $|a_0|$ value. The convergence characteristics of this version of the AH method for starting points $I_1 = (0.01,0.01)$, $I_2 = (0.6,0.6)$, and $I_3 = (1.5,0.5)$ are displayed on Figure 1. The I_1 point is close to the minimum point A of function (24), the first eigenvector $(\mathbf{u}_1)$ of its Hessian at this point is directed along OY axis, and the second one $(\mathbf{u}_2)$, along OX axis. In accordance with prescriptions of the third section, the displacement from I_1 is performed along $\mathbf{u}_1$ when a saddle point is required and along $\mathbf{u}_2$ when the maximum point is to be reached. The analogous situation takes place for starting points I_2 and I_3.

Conclusion

When we minimize a function using a certain descent method, there exists a simple criterion for the characterization of feasible directions of movement: the function values should decrease in subsequent points along the minimization path. Unfortunately, the situation is more complicated in the case when stationary points of nonzero indices are required. The most unpleasant is the case when the current point index differs from the required type. In the AH method, even in such a complicated situation, we can try to formulate a criterion for a feasible displacement choice. Indeed, due to the "betweenness" theorem (see the second section), if the displacement $\Delta \mathbf{x}$ results in the decreasing of λ_{k+1}, then all the eigenvalues $\varepsilon_1, \ldots, \varepsilon_k$ undergo nonpositive shifts, and if displacement results in increasing of λ_{k+1}, then all the eigenvalues $\varepsilon_{k+1}, \ldots, \varepsilon_n$ undergo nonnegative shifts. Therefore, if $\varepsilon_k < \varepsilon_{k+1} < 0$, then feasible displacement should result in increasing of the λ_{k+1} value, and if $0 < \varepsilon_k < \varepsilon_{k+1}$, then, on the contrary, feasible displacement should lead to its decreasing. Unfortunately, this criterion requires, in the general case, the knowledge of the third derivatives of the object function. In the present article, we have confined ourselves to a simpler question: When is displacement along a certain augmented Hessian eigenvector feasible?

TABLE I _____

Three versions of the AH method: example of convergence to stationary points of different types on the Cerjan–Miller surface with $a = c = 1$ and $b = 1.5$.

Starting point coordinates	Hessian eigenvalues $(\varepsilon_1, \varepsilon_2)$ calculated at the starting point	Stationary point to be reached	No. iterations		
			First version	Second version	Third version
$(0.01, 0.01)^a$	$(0.999, 1.999)$	A	1	1	1
		B	3	3	6
		D	3	3	5
$(0.6, 0.6)^a$	$(-1.059, 0.959)$	A	4	4	4
		B	3	3	3
		D	3	3	3
$(1.5, 1.5)^a$	$(-0.476, 0.748)$	C	3	3	3
		D	4	4	4
$(0.2, 0.2)$	$(0.808, 1.527)$	A	2	2	2
		B	4	4	4
		D	3	3	4
$(0.8, 0.1)$	$(-0.007, -1.440)$	A	4	4	5
		B	3	3	5
		D	3	3	4
$(1.5, 1.0)$	$(-1.043, 1.345)$	C	3	3	3
$(0.79, 0.79)$	$(-1.095, 0.924)$	A	5	5	5
		B	2	2	2
		D	5	5	5
$(0.99, 0.01)$	$(-1.485, -0.103)$	A	4	4	4
		C	5	6	5
		D	1	1	2

[a]Convergence paths from this point are displayed on Figure 1.

We hope that our prescriptions for the displacement vector choice will be of use for MCSCF optimization, transition-state search, etc.

References

1. C. M. Smith, Int. J. Quantum. Chem. **37**, 773 (1990).
2. J. Nichols, H. Taylor, P. Schmidt, and J. Simons, J. Chem. Phys. **92**, 340 (1990).
3. T. Helgaker, Chem. Phys. Lett. **182**, 503 (1991).
4. D. J. Wales, J. Chem. Soc., Faraday Trans. **88**, 653 (1992).
5. D. A. Liotard, Int. J. Quantum Chem. **44**, 723 (1992).
6. B. H. Lengsfield III, J. Chem. Phys. **73**, 382 (1980).
7. D. R. Yarkony, Chem. Phys. Lett. **77**, 634 (1981).
8. B. H. Lengsfield III and B. Liu, J. Chem. Phys. **75**, 478 (1981).
9. R. Shepard, I. Shavitt, and J. Simons, J. Chem. Phys. **76**, 543 (1982).
10. P. Jorgensen, P. Swanstrom, and D. L. Yeager, J. Chem. Phys. **78**, 347 (1983).
11. J. T. Golab, D. L. Yeager, and P. Jorgensen, Chem. Phys. **78**, 175 (1983).
12. H. J. Aa. Jensen and P. Jorgensen, J. Chem. Phys. **80**, 1204 (1984).
13. A. Banerjee, N. Adams, J. Simons, and R. Shepard, J. Phys. Chem. **89**, 52 (1985).
14. H.-J. Werner, in *Ab Initio Methods in Quantum Chemistry*, K. P. Lowley, Ed. (Wiley, New York, 1987), p. 1.
15. R. Shepard, in *Ab Initio Methods in Quantum Chemistry*, K. P. Lowley, Ed. (Wiley, New York, 1987), p. 63.
16. H. J. Aa. Jensen, H. Agren, and J. Olsen, in *Modern Techniques in Computational Chemistry: MOTECC-90*, E. Clementi, Ed. (ESCMOM, Leiden, 1990), p. 435.
17. D. Poppinger, Chem. Phys. Lett. **35**, 550 (1975).
18. D. J. Wales, J. Chem. Soc., Faraday Trans. **86**, 3505 (1990).
19. Yu. G. Khait and A. I. Panin, Int. J. Quantum Chem. **43**, 827 (1992).
20. C. J. Cerjan and W. H. Miller, J. Chem. Phys. **75**, 2800 (1981).
21. B. N. Parlett, in *The Symmetric Eigenvalue Problem* (Prentice-Hall, Englewood Cliffs, NJ, 1980).
22. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Addison Wesley, Reading, MA, 1958).
23. R. Fletcher, in *Practical Methods of Optimization* (Wiley, New York, 1980), Vol. 1.
24. J. Simons, P. Jorgensen, H. Taylor, and J. Ozment, J. Phys. Chem. **87**, 2745 (1983).