Strategies for Walking on Potential Energy Surfaces Using Local Quadratic Approximations

JACK SIMONS and JEFF NICHOLS*
Chemistry Department, University of Utah, Salt Lake City, Utah 84112

Abstract

An algorithm for locating stationary points corresponding to local minima and transition states on potential energy surfaces is further analyzed. This method utilizes local gradient and Hessian (i.e., first and second energy derivative) information to generate a series of "steps" that are followed to the desired stationary point. By designing the step sequence to move energetically downhill in all coordinates, local minima can be found. By stepping uphill along one local eigenmode of the Hessian while minimizing the energy along all other modes, one locates transition states. Key elements of this development are more efficient parameterization of the step vector in terms of quantities that permit the direction (i.e., uphill or downhill), and length of the step to be carefully controlled, and implementation of the ability to explore "side channels" as attractive options occur.

Introduction

Locating stationary points on energy surfaces, given knowledge of the local gradients and curvatures, represents a challenging and important problem in computational chemistry [1]. Such points correspond to geometries at which all gradients (first derivatives with respect to coordinates) vanish. They include minima, where all eigenvalues of the second derivative or Hessian matrix are positive, and transition states, where the Hessian has one negative eigenvalue. There are, of course, stationary points at which more than one Hessian matrix eigenvalue is negative; they correspond to "mountain tops" and are usually not of as much importance in chemistry. An algorithm that efficiently locates the desired stationary points would be of great utility.

In earlier publications [2a–2d,2h], we described the development and implementation of such an algorithm. It is the purpose of this work to build upon this earlier work and provide enhancements that we have found to produce improved behavior.

In this article, we describe a procedure for locating minima and transition states and for walking in the streambeds connecting them. At each step, the method uses the local slope or gradient ($\mathbf{F}$) vector and curvature or Hessian matrix ($\mathbf{H}$) to compute a step vector ($\mathbf{x}$) which is added to the current atomic coordinates ($\mathbf{r}_o$) to obtain new coordinates ($\mathbf{r}$) at which new $\mathbf{F}$ and $\mathbf{H}$ matrices are computed so the process can be continued. The Hessian matrix $\mathbf{H}$ may be evaluated using full an-

* Permanent address: IBM Corporation, Utah Supercomputer Institute, Salt Lake City, Utah 84112.

International Journal of Quantum Chemistry: Quantum Chemistry Symposium 24, 263–276 (1990)

CCC 0020-7608/90/010263-14$04.00

alytical energy derivative information, or may be computed via update or finite-difference methods; the approach described here can be used in all of these cases.

Convergence to a desired stationary point is reached when the norm of **F** is less than some specified tolerance, the number of negative eigenvalues of **H** is correct (i.e., zero for a minimum and one for a transition state), and the energy change from step to step E–E_o is within some tolerance.

We include in the class of problems to be examined "walks" that lead to fragmentation. For such events, a true transition state may not be reached because the potential energy surface only asymptotically approaches a point at which the forces vanish. Moreover, upon fragmentation, additional zero eigenvalues appear in the Hessian matrix* corresponding to the new translations and rotations that exist in the fragments but were internal modes in the original molecule.

The Method

The Local Quadratic Approximation

We begin by writing a local quadratic approximation to the potential energy surface in terms of the $3N$ cartesian components of the gradient, Hessian, and step matrices:

$$E = E_o + \mathbf{xF} + 1/2\mathbf{xHx} \, . \qquad (1)$$

By assumption, we know the energy surface only locally. Therefore, it is important to constrain our steps **x** to lie *within* a radius L for which the quadratic representation in Eq. (1) is valid. The determination of this "trust radius" L in terms of the ability of Eq. (1) to predict energy changes experienced for steps **x** within L is dealt with in a later section. In essence, the largest step length permitted is dynamically determined by comparing the energy change predicted by the local quadratic approximation and the actual energy change experienced when the predicted step **x** is taken.

Partitioning Into Internal and External Degrees of Freedom

If the energy surface $E(\mathbf{x})$ pertains to a molecule in the gas phase for which translational and rotational motions have no restoring forces, five or six coordinates can be removed to yield **F**, **H**, and **x** in the $3N$–5 or $3N$–6 internal coordinates (in practice, most *ab initio* electronic structure codes yield **F** and **H** in terms of $3N$ cartesian coordinates). We then partition these matrices into external and internal spaces by first constructing five or six orthonormal unit vectors that span the translational and (infinitesimal) rotational [2d] spaces together with $3N$–5 or $3N$–6

* For example, in the $H_2CO \rightleftharpoons H_2 + CO$ reaction, the reaction path plus three other modes are formed that have zero gradients and zero Hessian eigenvalues. In such cases, the walking algorithm must properly identify such degrees of freedom and not be "confused" by their presence (e.g., numerical precision limitations may cause the Hessian eigenvalues corresponding to these modes to be small and of arbitrary sign rather than identically zero). In particular, it is important that artifactual small negative Hessian eigenvalues not be confused with the physically relevant negative eigenvalue belonging to a transition state.

other orthonormal vectors that span the internal space. Projection of **F** and **H** onto these spaces then provides *internal* gradient and Hessian matrices. The components of **F** and **H** lying within the translational and rotational spaces can be ignored when dealing with an isolated species for which $E(\mathbf{x})$ is a function only of the internal coordinates.

If there are degrees of freedom which, on the basis of symmetry, have zero gradient components, they can also be removed from immediate consideration. For example, when examining the $^1A_1$ state of the H_2O molecule in C_{2v} geometry, the component of **F** along the b_2-symmetry asymmetric distortion coordinate vanishes. The Hessian eigenmode belonging to this degree of freedom may be removed from the "active space" if one desires to preserve C_{2v} symmetry. At a later point in the potential surface walk, one can restore this degree of freedom to active consideration (an action that will become essential if the $^1A_1$ state of H_2O under study is crossed by a state of $^1B_2$ symmetry).

Analysis in the Hessian Eigenmode Basis

Given **F** and **H**, the local quadratic energy expression can be rewritten in terms of the eigenmodes of **H**:

$$\mathbf{H}\mathbf{v}_i = h_i\mathbf{v}_i\,, \qquad i = 1, 2, \cdots 3N\text{–}5 \quad \text{or} \quad 3N\text{–}6 \tag{2}$$

(or fewer if additional degrees of freedom have been eliminated) where it takes on the form

$$E = E_o + F_i x_i + 1/2 x_i^2 h_i\,. \tag{3}$$

Here the Einstein summation convention is used, h_i is the ith eigenvalue of the Hessian ($h_1 \leq h_2 \leq h_3 \cdots$), and x_i and F_i are the components of **x** and **F** along the ith eigenmode of **H**:

$$x_i = \langle \mathbf{x} | \mathbf{v}_i \rangle\,, \quad \text{and} \tag{4a}$$

$$F_i = \langle \mathbf{F} | \mathbf{v}_i \rangle\,. \tag{4b}$$

Parameterization of the Step Vector

A primary element of the algorithm described here is the introduction of steps $\{x_i\}$ parameterized as:

$$x_i = \alpha F_i(\lambda - h_i)^{-1}\,. \tag{5}$$

The origin of this particular form of the step vector is detailed in Ref. [2h]. Briefly, the choice $\lambda = 0$, $\alpha = 1.0$ corresponds to a Newton–Raphson (NR) step, which may be acceptable if it is not too long *and* if the local Hessian eigenvalues correspond to the stationary point being sought. For example, if all of the h_i are positive, the NR step is not appropriate if one is searching for a transition state, but may be acceptable if one is searching for a minimum, because taking the NR step moves one to an approximate stationary point where the Hessian eigenvalues retain their original values.

The introduction of values of λ and of α different from 0 and 1.0 can be understood as follows. When used in the above quadratic energy expression, the postulated step gives the following first-order, second-order, and total energy changes:

$$E - E_o = \alpha F_i^2(\lambda - h_i)^{-1} + 1/2\alpha^2 h_i F_i^2(\lambda - h_i)^{-2} \quad (6a)$$

$$= \alpha F_i^2(\lambda - h_i)^{-2}\{\lambda - h_i(1 - \alpha/2)\}. \quad (6b)$$

The parameter λ is used to permit the step along the ith mode to either be opposite in sign from the corresponding gradient (if $\lambda < h_i$) or directed along the gradient (if $\lambda > h_i$); this criterion relates to our desire to control whether the linear term $\alpha F_i^2(\lambda - h_i)^{-1}$ in the local quadratic prediction of the energy change increases (in which case the linear contribution $x_i F_i$ should be positive) or decreases (for which $x_i F_i$ should be negative) along a particular mode.

The sign of the second-order energy change $1/2\ h_i x_i^2$ is determined entirely by the sign of the Hessian eigenvalue h_i; for positive h_i values, it is an increasing function of x_i, and for negative h_i it is a decreasing function of x_i. The sign of the total (first plus second order) energy change along each mode is determined by the sign of $\lambda - h_i(1 - \alpha/2)$; if this quantity is positive, the local quadratic energy increases, otherwise it decreases. Choices for the α parameter different from 1.0 are made only in situations illustrated by the following case:

(i) Suppose that $h_1 > 0$, and that $h_2/2$ (which is also > 0) is less than h_1.
(ii) Further suppose that one wishes to step uphill along the first eigenmode while moving downhill along the second (and all higher) modes.
(iii) In this case, the desire to cause the linear and total energy along the first mode to increase dictates that λ be chosen to obey: $\lambda > h_1$ and $\lambda > h_1(1 - \alpha/2)$; the more restrictive of these two constraints is $\lambda > h_1$.
(iv) The desire to move downhill along the second and higher modes requires that: $\lambda < h_2$ and $\lambda < h_2(1 - \alpha/2)$; the more restrictive of these is $\lambda < h_2(1 - \alpha/2)$; the choice $\alpha = 1.0$ then gives $\lambda < 1/2\ h_2$.
(v) If $h_2/2$ lies below h_1, the conditions on λ imposed by the desire to move uphill along the first mode ($\lambda > h_1$) and downhill along all others ($\lambda < h_2/2$) are incompatible.
(vi) In this case, choosing the largest value of α for which there exists a nonvanishing region within which λ can be chosen gives $\alpha = 2(h_2 - h_1)/h_2$, which then implies that $\lambda = h_1$. This particular choice produces a step lying entirely along the first mode, and is thus unacceptable. In such cases, a smaller value of α must be chosen; we take $\alpha = (h_2 - h_1)/h_2$, which then gives $h_1 < \lambda < (h_1 + h_2)/2$ as a range within which λ can be selected to "guide" the walk.

In summary, the α parameter is used only when $h_2/2$ lies below h_1, and is used to provide a nonvanishing range of choices for λ that will permit the energy to move uphill or downhill as desired.

Problems With "Stitching"—Short Steps are Better Than Long Steps

In our earlier work [2a–2d,2h], emphasis was placed primarily on the total energy change along each mode, rather than insisting that both the linear and total predicted

energy changes be as desired. This point of view can lead to step lengths that "overshoot" the location of minima or are too long in the sense that they give displacements beyond the region where the local quadratic approximation is accurate.

To clarify these points, consider the contribution to the local quadratic energy surface along a particular mode when one is near a local minimum or transition state of interest:

$$\Delta E_j = F_j x_j + 1/2 x_j^2 h_j \,. \tag{7}$$

If h_j is positive, displacements x_j occur on an upward curved parabola; if h_j is negative, x_j is on a downward curved parabola. In either case, a specified desirable energy change (ΔE positive in the $h_j < 0$ case and ΔE negative in the $h_j > 0$ case) can be realized *either* by: (i) Taking a *small* step along which the linear energy change is of the desired sign while the quadratic term is small and of opposite sign (n.b., in the case considered here, the sign of the quadratic term is *always* opposite to what one wants; if $h_j > 0$, the quadratic term is positive, and one is trying to minimize the energy; if $h_j < 0$, the quadratic term is negative, and one is trying to maximize the energy), or by (ii) Taking a *larger* step in the same direction [i.e., with x_j having the same sign as used above in (i)] along which the linear energy change is much larger yet still of the desired sign while the quadratic contribution is large and of opposite sign.

We prefer to take the smaller steps characterized in (i) above. The longer steps of (ii) suffer two drawbacks: (1) Being longer, they are more likely to move the coordinates outside the region where the local quadratic approximation used to generate the step is valid (it is essential to keep in mind that the local quadratic surface is *not* the real surface). (2) They generate steps which "stitch"; that is, steps that move back and forth across the streambed along which the walk proceeds. In contrast, the steps of (i) are found to undergo little stitching.

The Form of the Step Elements

In our earlier work on this subject [2a–2d,2h], the form for the step elements shown in Eq. (5) was shown to arise from making the local quadratic energy functional of Eq. (1) stationary subject to the constraint (imposed by Lagrange multiplier λ) that the step be of a specified length. Here, instead of focusing on the constraint of the step length, we assume steps parameterized as in Eq. (5) and ask how λ and α can be determined to generate an optimal streambed walk, with λ chosen primarily to guide the *direction* of the step and α used to guarantee that the Hessian eigenvalue structure will permit a step of the desired character.

If a calculation then produces a *trial* step that is too long, in the sense that it yields an energy change that was not accurately predicted by the local quadratic approximation, this step can be further reduced in length. It is straightforward to see that doing so will not change the fact that the local quadratic approximation yields energy changes of the desired signs along all modes (because both the linear and total energies are required to be of the correct sign).

As an alternative to our strategy, one may introduce a procedure in which:

The choice $\alpha = 1$ causes $\lambda = 0$ to be a stationary point corresponding to a minimum ($d^2E/d\lambda^2$ is positive) if h_1 is positive. This means that a Newton–Raphson step [2a] (i.e., the step with $\lambda = 0$ and $\alpha = 1$) is optimal as long as $h_1 > 0$ and as long as this unscaled step is within the so-called trust radius L.

If the Newton–Raphson step is too long (i.e., if it exceeds L), any choice of $h_1 > \lambda > 0$ will also generate too long a step ($x_i = F_i(\lambda - h_i)^{-1}$) because the magnitude of $(\lambda - h_i)^{-1}$ will be larger than for $\lambda = 0$. Therefore, a value of $\lambda < 0$ must be chosen. In the range $\lambda < 0$ for $h_1 > 0$, $dE/d\lambda$ is negative; $d^2E/d\lambda^2$ is positive for $-h_1/2 < \lambda < 0$. Therefore, the most negative change in E will be realized for λ as large as possible. Choosing λ too close to zero generates an unscaled step length that is very large, as a result of which α must be chosen to reduce the step length, which then reduces the magnitude of $(E - E_o)$. In this case, the best choice of λ is the value that yields the maximum acceptable step length L. Therefore, we simply pick λ ($\lambda < 0$) by solving $F_i^2(\lambda - h_i)^{-2} = L^2$, after which α-scaling is unnecessary. As long as the force along the h_1 mode is nonnegligible, we can approximate the solution of this equation (which, in general, we solve iteratively) by $\lambda = h_1 - |F_1/L|$ because λ is closest to h_1 than to any other h_j. On the other hand, if F_1 is very small (e.g., if the h_1 mode is a symmetry-breaking mode or if one is near to a stationary point along h_1), an approximation to the solution of the above equation that is even more conservative can be written as $\lambda = h_1 - |F|/L$, where $|F|$ is the norm of the full gradient. With either choice of λ, a step that is approximately L in length is obtained (the step formed using this λ value is actually scaled subsequently to be exactly of length L).

If One is Not Near a Minimum ($h_1 < 0$). On the other hand, if h_1 is negative, the Newton–Raphson step $\lambda = 0$ is unacceptable because it violates the energy-lowering condition $\lambda < h_1$ determined earlier. In the range $\lambda < h_1$, $dE/d\lambda$ is negative and $d^2E/d\lambda^2$ is negative. Therefore, the most negative change in E will again be realized for λ as large as possible (i.e., as close to h_1 as possible). Once again, we pick λ by solving $F_1^2(\lambda - h_1)^{-2} = L^2$, after which α-scaling is unnecessary.

In summary, *to perform energy minimization walks,* we choose λ and α as follows:

(1) If $h_1 > 0$, we take the Newton–Raphson step ($\lambda = 0$) if its length is within L, and we set α equal to 1.0.
(2) If $h_1 > 0$, but the Newton–Raphson step length exceeds L, we determine $\lambda(\lambda < 0)$ by requiring $F_j^2(\lambda - h_j)^{-2} = L^2$ (or $|F|^2/(\lambda - h_1)^2 = L^2$ to be more conservative) to be obeyed. This yields $\lambda = h_1 - |F_1/L|$ (or, more conservatively, $\lambda = h_1 - |F|/L$). Again, we take $\alpha = 1.0$.
(3) If $h_1 < 0$, we determine $\lambda(\lambda < h_1)$ by requiring $F_j^2(\lambda - h_j)^{-2} = L^2$ to be obeyed ($\lambda = h_1 - |F_1/L|$) and α is once again set to 1.0.

In both (2) and (3), the step is formed with the specified value of λ, which yields a total step length near L (which can then be rescaled to be exactly of length L).

Transition-State Walks

Walks up the Lowest Eigenmode. Moving "uphill" along the lowest Hessian eigenmode while remaining at minima along the other eigenmodes generates a

(i) The local quadratic energy function is minimized along all modes but one using the Newton–Raphson or scaled NR step.
(ii) The energy is maximized along a single mode (the direction along which one wishes to move uphill).

The difficulties with implementing such an approach are:

(i) If the local Hessian eigenvalue belonging to the mode along which one is stepping uphill is positive, the task of maximizing the energy along this mode is ill defined; one needs to specify the maximum step length along the uphill mode, and to then take this maximum step.
(ii) Even once the Hessian eigenvalue along the mode one is following uphill becomes negative, one is faced with controlling the step length along this direction and *separately* along the other modes. In essence, one must introduce two maximum step lengths because one has separated the problem into two problems: minimization along all modes but one, and maximization along one.

This procedure is certainly one which makes sense; however, we have chosen, after considerable experimentation with both approaches, to follow the strategy detailed in this article.

Walks to Minima and to Transition States

Minimization Walks

As displayed in Eq. (6), steps for which both $\lambda < h_i$ and $\lambda < h_i(1 - \alpha/2)$ yield *negative linear* $(F_i x_i)$ *and negative total* energy changes along the ith mode. This is, of course, a property that a walk to a minimum along the ith mode should possess. If the lowest Hessian eigenvalue h_1 is positive (this is characteristic of regions near local minima) $\lambda < h_1(1 - \alpha/2)$ is the more restrictive constraint; a λ that obeys this will also obey $\lambda < h_i(1 - \alpha/2)$ for all other modes because the h_i are arranged in increasing order. If, on the other hand, h_1 is negative (this is characteristic of regions near transition states), $\lambda < h_1$ is the more restrictive constraint. Again, if λ obeys this condition, it will automatically also obey $\lambda < h_i$ and $\lambda < h_i(1 - \alpha/2)$ for the other modes, independent of whether the other h_i are positive or negative.

If One is Near a Minimum ($h_1 > 0$). Thus, for positive h_1, one wants $\lambda < h_1(1 - \alpha/2)$ and for negative h_1 one needs $\lambda < h_1$. Although these statements limit λ, they do not determine the optimal value of λ. To do so, we examine the dependence of the quadratic energy functional on λ for values of λ that obey the above conditions. Differentiation of the quadratic energy functional to seek a value of λ for which it is stationary yields:

$$dE/d\lambda = -\alpha F_i^2(\lambda - h_i)^{-3}\{\lambda - h_i(1 - \alpha)\}\,, \quad \text{and}$$
$$d^2E/d\lambda^2 = 2\alpha F_i^2(\lambda - h_i)^{-4}\{\lambda - h_i(1 - 3\alpha/2)\}\,.$$

"streambed walk" along this lowest mode. As explained earlier, steps for which $\lambda < h_2$ and $\lambda < h_2(1 - \alpha/2)$ produce linear and total energy lowering along all modes other than the first. To generate linear and total energy *increases* along the h_1 mode requires that $\lambda > h_1$ and $\lambda > h_1(1 - \alpha/2)$.

If One is Not Near a Transition State ($h_1 > 0$). If h_1 and h_2 are both positive (which is characteristic of geometries near local minima), these constraints reduce to $h_1 < \lambda < h_2(1 - \alpha/2)$. If $h_2/2$ is less than h_1, the choice $\alpha = 1$ (which corresponds to an unscaled step) can not be used because no value of λ obeys $h_1 < \lambda < h_2/2$. It is in such circumstances that the α-parameter is used; it allows us to generate a nonzero range for selecting the λ-parameter to achieve a step with the desired characteristics.

The largest value of α (i.e., the least scaling of the step) that gives rise to a nonzero range for choosing λ is $\alpha = 2(h_2 - h_1)/h_2$. This particular α gives $h_2(1 - \alpha/2) = h_1$, as a result of which λ is bounded to $h_1 < \lambda < h_1$. The choice $\lambda = h_1$ generates a step lying entirely along the h_1 eigenmode; this step has no ability to incorporate movement along the other modes, and is therefore unacceptable (because unless all $F_j = 0$ for $j > 1$, these other modes are not at their optimal positions). To maintain a distinct range within which λ can be chosen, we choose to take an even smaller step and select $\alpha = (h_2 - h_1)h_2$ as the scaling parameter. This then restricts λ to the range $h_1 < \lambda < (h_1 + h_2)/2$, within which we choose the midpoint $\lambda = (h_1 + (h_1 + h_2)/2)/2$. Using these values of α and λ, the step $x_i = \alpha F_i(\lambda - h_i)^{-1}$ is evaluated. If the length of this $\mathbf{x}$ is less than L, it is taken; if the step length exceeds L, x_i is further scaled back to yield a total step length equal to L.

If, in contrast to the above situation, $h_2/2$ exceeds h_1, the choice $\alpha = 1.0$ is acceptable, in which case we choose λ as the midpoint of the two bounds: $\lambda = (h_1 + h_2/2)/2$. Again, if the step x_i obtained using this λ value and $\alpha = 1.0$ exceeds L, it is further scaled back to L.

If One is Near a Transition State ($h_1 < 0$). If h_1 is negative, the appropriate constraints $h_1(1 - \alpha/2) < \lambda < h_2(1 - \alpha/2)$ can be met with $\alpha = 1$. In this case, we again choose λ as the midpoint of this range: $\lambda = (h_2 + h_1)/4$, and we further scale back the x_i if the total step length obtained with this λ,α combination exceeds L.

In summary, *to walk uphill* along the streambed belonging to the *lowest Hessian eigenvalue,* we:

(1) Take $\alpha = 1.0$ and $\lambda = (h_1 + h_2/2)/2$ if h_1 is positive and $h_2/2$ exceeds h_1.
(2) Take $\alpha = (h_2 - h_1)/h_2$ and $\lambda = (h_1 + (h_1 + h_2)/2)/2$ if h_1 is positive but $h_2/2$ does not exceed h_1.
(3) Take $\alpha = 1.0$ and $\lambda = (h_2 + h_1)/4$ if h_1 is negative.

In all three cases, the step elements $x_i = \alpha F_i(\lambda - h_i)^{-1}$ are further scaled back if their total length exceeds L. A step obtained by scaling back a step that has linear and total energy changes of the desired sign will also have linear and total energy changes of the proper sign (because the quadratic energy change is always of the "wrong" sign, and cutting x_j back reduces the magnitude of $1/2\ h_j x_j^2$ more than that of $F_j x_j$).

Walks up Other Eigenmodes.

The Extra Difficulties for Modes Other Than the Lowest. To walk uphill along a streambed that belongs to one of the other Hessian eigenmodes requires additional care. Choosing λ to obey $\lambda > h_2$ and $\lambda > h_2(1 - \alpha/2)$ will certainly generate a step that, within the local quadratic approximation, has positive linear and total energy changes along the h_2 mode. However, this choice will also cause the step to increase in energy along the h_1 mode. Such behavior is not characteristic of the desired streambed walk; the energy is supposed to be minimized for all modes except the one (h_2) along which uphill movement occurs.

Coordinate Scaling to "Stretch" the Potential Energy Surface. The key to being able to walk uphill along eigenmodes that are not the lowest using the step vector parameterization advocated here is to "*scale*" *the coordinate* along the mode of interest to "stretch" the quadratic energy surface along this direction. By so doing, one makes this direction appear to have the smallest curvature.

In particular, we express the step component along the mode (h_k in general) to be followed as:

$$x_k = \beta y_k \, .$$

In terms of step components $\{x_1, x_2, \cdots x_{k-1}, y_k, x_{k+1}, \cdots x_{3N-6 \text{ or } 3N-5}\}$, the local quadratic energy functional becomes:

$$E = E_o + F_i x_i + 1/2 h_i x_i^2 + (F_k \beta) y_k + 1/2 (h_k \beta^2) y_k^2 \, ,$$

where the sum over i runs over $1, 2, \cdots k-1, k+1, \cdots 3N-6$ or $3N-5$.

When viewed as a function of the step components $\{x_1, x_2, \cdots x_{k-1}, y_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$, this local quadratic energy surface appears to have a Hessian eigenvalue of $\beta^2 h_k$ where the original surface had h_k, and to have a gradient βF_k where the original surface had F_k.

By choosing β such that $\beta^2 h_k$ lies below $h_1/2$ (so that the new lowest Hessian eigenvalue $\beta^2 h_k$ is less than $1/2$ the new second lowest eigenvalue so that λ can be chosen in the range $\beta^2 h_k < \lambda < h_1/2$), one can then employ the transition-state walking strategy appropriate to the lowest eigenmode direction as detailed above. In each such step, however, it is essential to keep in mind that one is generating step components $\{x_1, x_2, \cdots x_{k-1}, y_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$; the y_k component must then be multiplied by β to obtain x_k. Once the $\{x_1, x_2, \cdots x_{k-1}, x_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$ are in hand, transformation to cartesian or to internal coordinate displacements can be performed.

When implementing the above coordinate-scaling method for walking up higher eigenmodes, it is important to understand that the maximum step length L appropriate for the walk in the $\{x_1, x_2, \cdots x_{k-1}, y_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$ space is not necessarily the same as that for the $\{x_1, x_2, \cdots x_{k-1}, x_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$ space. This can be understood by considering the length of the $\{x_1, x_2, \cdots x_{k-1}, y_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$ vector and of the resulting coordinate step vector. If $\{x_1, x_2,$

$\cdots x_{k-1}, y_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$ is constructed to be of length L', then $\{x_1, x_2, \cdots x_{k-1}, x_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$ will be of length L, where

$$(L)^2 = (L')^2 + (\beta^2 - 1)y_k^2 .$$

Because the step $\{x_1, x_2, \cdots x_{k-1}, y_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$ is constructed to have a significant or even dominant y_k component, y_k^2 will often be close to $(L')^2$, in which case $(L)^2$ will approach $\beta^2(L')^2$. Therefore, to achieve a coordinate step $\{x_1, x_2, \cdots x_{k-1}, x_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$ within L, one should restrict the scaled-step $\{x_1, x_2, \cdots x_{k-1}, y_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$ to lie within $L' = \beta^{-1}L > L$. In practice, we do not so expand the maximum step length thus opting for a more "conservative" walk (as a result, our steps in $\{x_1, x_2, \cdots x_{k-1}, x_k, \cdots x_{3N-6 \text{ or } 3N-5}\}$-space are usually short).

Perspective

Thus far, we have covered the following aspects of our walking strategy:

(1) How to walk to local minima using steps that lie within a specified length L.
(2) How to walk uphill along the softest (h_1) eigenmode in search of a transition state, again using steps that lie within L.
(3) How to use coordinate scaling to distort the local quadratic energy surface so that a higher eigenmode becomes the lowest, thereby allowing one to walk uphill along higher eigenmodes using the same strategy.

What to Do When Eigenmodes Cross

When carrying out a walk directed toward a local minimum, the λ-parameter is always chosen less than the lowest Hessian eigenvalue or less than one-half of this eigenvalue. This condition remains in effect even if the eigenvalues undergo crossings as one steps along the potential surface. Hence, eigenvalue crossings do not adversely affect the behavior of walks to minima.

In contrast, when walking toward a transition state, such crossings give rise to qualitatively important changes in the step if the crossing involves the mode (e.g., h_k) along which uphill movement has been directed. For example, when following h_1 uphill, h_2 may decrease (or h_1 increase) until the two cross. Whenever such crossings occur, a *decision must be made.* One can choose to follow the physical direction of the streambed along which one had been stepping, or one can begin to move along the eigenmode that has just crossed. Neither decision is right or wrong; both represent reasonable choices, and either (or neither) may lead to a transition state. In principle, both should eventually be "explored." It is the purpose of this section to specify how one can effect these two choices.

Tracking the Original Streambed. To continue along the *direction* that the walk had been following prior to the eigenmode crossing, one must implement a so-called "**tracking**" method. We use an eigenvector tracking method in which the scalar product of the Hessian eigenvector (v_k) corresponding to the "uphill" mode is computed from step to step. The quantity $\langle v_k(\text{step } n) | v_k(\text{step } n+1) \rangle$ should

be approximately 1.0 at each step. This allows the desired eigenvector v_k(step $n+1$) to be properly identified even as its eigenvalue is crossed by other eigenvalues.

Of course, if one desires to track as described above, it is important to realize that one is now following an eigenmode whose position in the order $h_1 < h_2 < h_3 < \cdots$ has changed. Consider for example, the situation in which one had been following and wishes to continue to track the physical direction of the h_1 eigenmode uphill even after the h_2 mode crosses and moves below the previous h_1 mode. In this case, the above scalar product technique can identify the eigenvalue crossing and specify that one should now follow the *new* h_2 direction. To do so, the coordinate scaling device described earlier must be invoked. Therefore, to properly track a specified direction one has to use the tracking and coordinate scaling devices in concert.

Moving to Another Mode by Not Tracking. When the Hessian eigenvalue belonging to the mode being followed is crossed by another, it is possible to move from the streambed that one was pursuing to explore the direction of the eigenvalue that has just crossed. In the example considered in the preceding paragraph, after the h_2 mode moves below and crosses the h_1 mode, one can choose to not track on the *direction* specified bythe previous h_1 mode, but to follow uphill the new lowest mode (the previous h_2 direction which is now the h_1 direction). Such a decision is effected by not invoking a tracking device and simply allowing the walking algorithm to continue stepping uphill along the h_1 direction while keeping to minima along directions other than h_1.

Summary

This completes our description of how the stepping algorithm decides what **direction** to move in, given a maximum step length L within which the step is constrained. The procedure generates a step that may be taken to generate the next position about which a new local quadratic approximation to the energy surface will be formed. However, *there are circumstances under which the step put forth for consideration must not be taken* but, rather, replaced by an alternative (shorter) step. That is, there are circumstances under which the maximum step length L must be shortened. It is this aspect of the algorithm to which we now turn attention.

Step Length Control

The Maximum Step Size L

In the algorithm outlined in the previous section, each step is constructed to have a total length less than or equal to a preassigned maximum length L. The choice of L is very much a matter of taste and of "common sense." We prefer to generate walks that smoothly trace out the locus of points characterizing streambeds; therefore, we usually choose rather conservative L values (e.g., L less than 0.05 Å per cartesian displacement in the molecule). Clearly, L must be less than the dynamic range over which the true potential energy surface changes its features by amounts that are deemed important. For chemical bonds, changes of a few tenths of an Å usually correspond to appreciable energy changes.

Choosing L small requires many steps in the walk, thereby increasing the computational expense. Choices of L that are large are less harmful than they might seem at first glance because of the step-size reduction strategy detailed below. In a nutshell, any step for which the true energy realized at the displaced geometry is in sharp disagreement with the local quadratic prediction of the energy must undergo further step-size reduction. Of course, this reduction process entails computational expense, so one would like to use an L value that would not often necessitate such action. In essence, L should be chosen with a good deal of common sense.

Step-Size Reduction

The procedure for generating step sequences described in the previous section may produce a step that moves beyond the region where the local quadratic approximation to the true potential energy surface is valid. In such a case, the step must be further reduced until it lies within this range. Since we do not *a priori* know the true energy surface except at the point around which the local quadratic expansion is carried out (where the energy is E_o), we must allow the step generated by the algorithm detailed in the last section to be taken (on a trial basis) so that the true (E_T) energy at $\mathbf{r} = \mathbf{r}_o + \mathbf{x}$ can be evaluated. If the quadratic prediction

$$E - E_o = F_i x_i + 1/2 h_i x_i^2$$

accurately reproduces the true energy difference $E_T - E_o$ then the step can confidently be taken. On the other hand, if $F_i x_i + 1/2\ h_i x_i^2$ does not agree well with $E_T - E_o$ then the step $\{x_i\}$ must be reduced in length.

It remains to state what it means for the two energies to agree well. In our implementation, we insist that the predicted energy difference $E - E_o = F_i x_i + 1/2\ h_i x_i^2$ and the energy difference $E_T - E_o$ observed once the step is taken (on a trial basis):

(1) *Be of the same sign*—we do not want the predicted energy to direct the walk uphill only to find that the step actually moves downhill (this is indicative of a step for which the quadratic energy change, which is always undesirable in sign, has overcome the favorable linear term),

(2) *Be equal within some range* in the sense that $\text{Min}(|E - E_o|, |E_T - E_o|)/\text{Max}(|E - E_o|, |E_T - E_o|)$ be equal to unity within a specified tolerance.

In this most straightforward implementation of the concept, if agreement between $E_T - E_o$ and $E - E_o$ is not met, the L value is set equal to one-half the length of the step just tested, and a new step is computed (as above) using this smaller L value. This process is continued until the local quadratic approximation to the true surface is valid (in the sense described above) at which time the step is taken. It should be noted that generating such a series of step reductions does not require the evaluation of new gradient and Hessian matrices; the most time consuming element is the evaluation of E_T at each of the "trial steps."

Once a step to a new geometry is realized and new **F** and **H** matrices are computed, the stepping algorithm begins again. Each successive step is restricted by this algorithm to:

(1) Have a total length less than some specified maximum step size L, and
(2) Have its length further reduced to guarantee that the quadratic energy change accurately represents the true (observed) energy change for that step.

Overview and Summary

In this article, we have presented an algorithm that permits the location and characterization (via the nature of the Hessian eigenvalues) of local minima and transition states on potential energy surfaces. This method has the following characteristic features:

(1) It uses local gradient and Hessian information.
(2) It generates steps that produce the desired behavior (i.e., uphill or downhill) in *both* the linear and total quadratic energy changes along each Hessian eigenmode.
(3) It permits rotations and translations (and non-symmetry-preserving motions) to be removed from consideration.
(4) Through use of a maximum step size and a step-reduction strategy, it controls the step length to keep each step within a region where the local quadratic energy approximation is valid.
(5) It controls the step direction in a manner that is guaranteed to move either downhill in all Hessian eigenmode directions (when searching for minima) or uphill along one eigenmode and downhill along all others (when searching for transition states).
(6) It permits streambeds along any eigenmode of the Hessian to be explored by introducing a coordinate scaling device.
(7) It allows one to "track" on a particular eigenmode direction and follow it even if this eigenvalue is crossed by other eigenvalues during the walk (i.e., even if directions transverse to the stepping direction acquire smaller uphill curvature as the walk proceeds).
(8) Alternatively, it permits one to follow a particular eigenvalue's direction throughout a walk; as other eigenvalues move relative to the mode being followed, one can adjust and "switch" to the new mode. This strategy allows one to move to a new direction if one of smaller uphill curvature appears along the walk.

This walking algorithm is implemented and routinely used in our highly modular Utah MESS KIT electronic structures codes. Its primary elements can be summarized as follows (with the step components in the Hessian eigenmode basis given as $x_j = \alpha F_j/(\lambda - h_j)$):

To perform energy minimization walks, we choose λ and α as follows:

(1) If $h_1 > 0$, we take the Newton–Raphson step ($\lambda = 0$) if its length is within L, and we set α equal to 1.0.
(2) If $h_1 > 0$, but the Newton–Raphson step length exceeds L, we determine $\lambda(\lambda < 0)$ by requiring $F_j^2(\lambda - h_j)^{-2} = L^2$ (or $|F|^2/(\lambda - h_1)^2 = L^2$ to be more conservative) to be obeyed. This yields $\lambda = h_1 - |F_1/L|$ (or, more conservatively, $\lambda = h_1 - |F|/L$). Again, we take $\alpha = 1.0$.

(3) If $h_1 < 0$, we determine $\lambda(\lambda < h_1)$ by requiring $F_j^2(\lambda - h_j)^{-2} = L^2$ to be obeyed ($\lambda = h_1 - |F_1/L|$) and α is once again set to 1.0.

To walk uphill along the streambed belonging to the (current) lowest Hessian eigenvalue, we:

(1) Take $\alpha = 1.0$ and $\lambda = (h_1 + h_2/2)/2$ if h_1 is positive and $h_2/2$ exceeds h_1.
(2) Take $\alpha = (h_2 - h_1)/h_2$ and $\lambda = (h_1 + (h_1 + h_2)/2)/2$ if h_1 is positive but $h_2/2$ does not exceed h_1.
(3) Take $\alpha = 1.0$ and $\lambda = (h_2 + h_1)/4$ if h_1 is negative.

To walk uphill along a streambed belonging to another (current) Hessian eigenvalue (say h_k), we:

(1) Scale h_k by β^2 to produce an effective eigenvalue of $\beta^2 h_k$, and scale F_k by β;
(2) Choose β such that $\beta^2 h_k$ lies below $h_1/2$;
(3) Use the conventional walking algorithm but with the kth mode scaled as described here. After computing a step (y_k) with the scaling of the kth mode operative, the true step along x_k is computed as $x_k = \beta y_k$.

To continue to follow a particular direction even as other eigenvalues cross the eigenvalue corresponding to the direction being pursued, we invoke the eigenvector tracking device involving the quantity $\langle v_k(\text{step } n) | v_k(\text{step } n + 1) \rangle$.

Alternatively, we can choose not to track on the direction f the eigenmode that had been followed by simply noting when another eigenvalue crosses the eigenvalue of the mode that had been pursued and switching to the new mode at that time.

Acknowledgments

This work was supported in part by the Office of Naval Research and by NSF Grant No. 8814765.

Bibliography

[1] P. Pulay, *Adv. in Chem. Phys. LXIX,* K. P. Lawley, Ed. (John Wiley and Sons, New York, 1987), p. 241. F. Bernardi and M. A. Robb; and H. B. Schlegel, *Adv. in Chem. Phys. LXVII,* K. P. Lawley, Ed. (John Wiley and Sons, New York, 1987), p. 155, 249.

[2] (a) J. Simons, P. Jorgensen, H. Taylor, and J. Ozment, J. Phys. Chem. **87**, 2745 (1983); (b) D. O'Neal, H. Taylor, and J. Simons, J. Phys. Chem. **88**, 1510 (1984); (c) A. Banerjee, N. Adams, J. Simons, and R. Shepard, J. Phys. Chem. **89**, 52 (1985); (d) H. Taylor and J. Simons, J. Phys. Chem. **89**, 684 (1985); (e) C. J. Cerjan and W. H. Miller, J. Chem. Phys., **75**, 2800 (1981); (f) J. Baker, J. Comp. Chem. **9(5)**, 465 (1988); (g) J. Baker, J. Comp. Chem. **7(4)**, 385 (1986); (h) J. Nichols, H. Taylor, P. Schmidt, and J. Simons, J. Chem. Phys. **92**, 340 (1990).

Received April 18, 1990