CHAPTER 8

GEOMETRY OPTIMIZATION ON POTENTIAL ENERGY SURFACES

H. Bernhard Schlegel

Department of Chemistry, Wayne State University
Detroit, Michigan 48202, USA

Contents

1. Introduction

Geometry optimization is a key step in all theoretical studies of molecular structure and reactivity. With the advances that have been achieved in analytical energy derivatives over the last 15 years, finding equilibrium geometries has become almost routine at many levels of *ab initio* molecular orbital theory. This chapter focuses on energy derivative based optimization methods and deals with algorithms for locating features such as minima, transition states and reaction paths.

Analytical first derivatives or gradients are available for all but the most highly correlated levels of *ab initio* molecular theory.[1–4] In general, the computational time for gradients is similar to or less than the time required for the energy evaluation, making gradient calculations highly cost-effective for gathering information about potential energy surfaces for polyatomic molecules. Analytical second derivatives are more difficult to calculate but are routinely available for Hartree–Fock, MP2 and MCSCF theories as well as a few other levels of theory. Hence, optimization methods that rely only on gradients will be more widely applicable than algorithms that require frequent calculation of second derivatives. Methods for calculating energy derivatives are examined thoroughly in other chapters of this volume[1] and elsewhere.[2–4]

Algorithms for optimization have been discussed extensively in the numerical analysis literature.[5] Methods for finding equilibrium geometries and locating transition states by *ab initio* molecular orbital theory have been the subject of a number of past reviews.[12] The present chapter starts with a few definitions of features on potential energy surfaces (Sec. 2), and then proceeds to discuss of some of the more widely used methods for optimizing equilibrium geometries (Sec. 3), searching for transition states (Sec. 4), locating minima on seams of intersection and finding points of closest approach between two surfaces (Sec. 5), following reaction paths (Sec. 6) and finishes by considering some of the practical aspects of geometry optimization (Sec. 7). Because of space limitations and the range of material that needs to be covered, an attempt is made to keep the presentation concise. More detailed expositions on a number of aspects of geometry optimization can be found in the primary literature.

2. Concepts

2.1. *Potential Energy Surfaces*

The *potential energy surface* for a molecule describes the energy of the molecule as a function of the positions of the atoms or nuclei. To construct a potential energy surface, one must first separate the treatment of the electronic and nuclear motions, which can be done via the Born–Oppenheimer approximation. Because the electrons are so much lighter than the nuclei, the electronic wave function quickly re-adjusts to any nuclear motion. In the Born–Oppenheimer approximation, a potential energy surface is obtained by calculating the electronic energy and wave function for a series of fixed nuclear positions. This is usually very good for equilibrium geometries and most transition states, but is problematic near seams, conical intersections and weakly avoided crossings (see Sec. 5 and Refs. 19–22).

The concept of a potential energy surface allows one to talk about chemical structures. The *equilibrium geometry* of each isomer of a molecule corresponds to a valley or minimum on the potential energy surface. The position of the minimum yields the bond lengths and angles of the molecule. The concept of molecular structure is less clear when the minimum is very broad or shallow. A reaction is a movement from the valley representing the reactants to the valley of the products. The path across the landscape, including any intermediates and transition states, describes the *reaction mechanism.* The highest point along the lowest energy reaction path between reactants and products is the *transition state.* The transition state, in turn, governs the rate of the reaction. Thus, to use potential energy surfaces to discuss molecular structure and reactivity, one must be able to locate minima, transition states and reaction paths. To compute enthalpies of reaction by statistical mechanical methods,[23] one requires the shape of the potential energy surface at the minima (i.e. the vibrational frequencies). If transition state theory (or variational transition state theory) is used to estimate reaction rates,[24] one also needs the shape of the potential energy surface at the transition state (or for a short distance along the reaction path on either side of the transition state).

A variety of processes, such as photochemical reactions[21] and electron transfer reactions,[22] involve more than one potential energy surface. Rates of these reactions depend on the interaction between two surfaces and must be treated by theoretical methods that go beyond the Born–Oppenheimer approximation. Nevertheless, these treatments start with potential energy surfaces, either diabatic or adiabatic, and require information about the lowest point on the seam of crossing between the two surfaces or about the point of closest approach between the two surfaces. Optimization methods needed to find these points are discussed in Sec. 5.

2.2. *Minima, Transition States and Higher-Order Saddle Points*

The first derivatives of the potential energy surface are known as the *gradients.* The *forces* on the atoms or nuclei in a molecule are equal to the negative of the gradients. Since the forces or gradients are zero at minima, transition states and higher order saddle points, these are also called *stationary points.* Using only the first derivatives, one cannot distinguish between a minimum, a transition state and a higher order saddle point; one must examine the second derivatives. The matrix of second derivatives of the potential energy surface is also called the *Hessian matrix* or the *force constant matrix.* If the Hessian or force constant matrix is transformed to mass-weighted coordinates and diagonalized, then the eigenvectors are the normal modes of vibration and the eigenvalues are proportional to the squares of the vibrational frequencies. The nature of a stationary point (whether it is a minimum, transition structure, second-order saddle point, etc.) is invariant to coordinate transformations.

For the potential energy surface to have a minimum with respect to a particular coordinate, the second derivative of the surface with respect to that coordinate must be positive; to be a maximum, the second derivative must be negative. In order that a stationary point describe an equilibrium geometry, that point must be a minimum with respect to all directions. Since the coordinates of a molecule can interact strongly, it is best to diagonalize the Hessian or force constant matrix first. The eigenvectors form a set of orthogonal, non-interacting directions, and a minimum is required to have positive eigenvalues for all modes. If the potential energy surface permits translation and/or rotation of the molecule, then the eigenvalues with respect to these modes should be zero and these motions should

be projected out. Equivalently, for a molecule to be at a minimum, all of its vibrational frequencies must be real (i.e. no imaginary frequencies). Figure 1 shows a number of minima on a model potential energy surface.

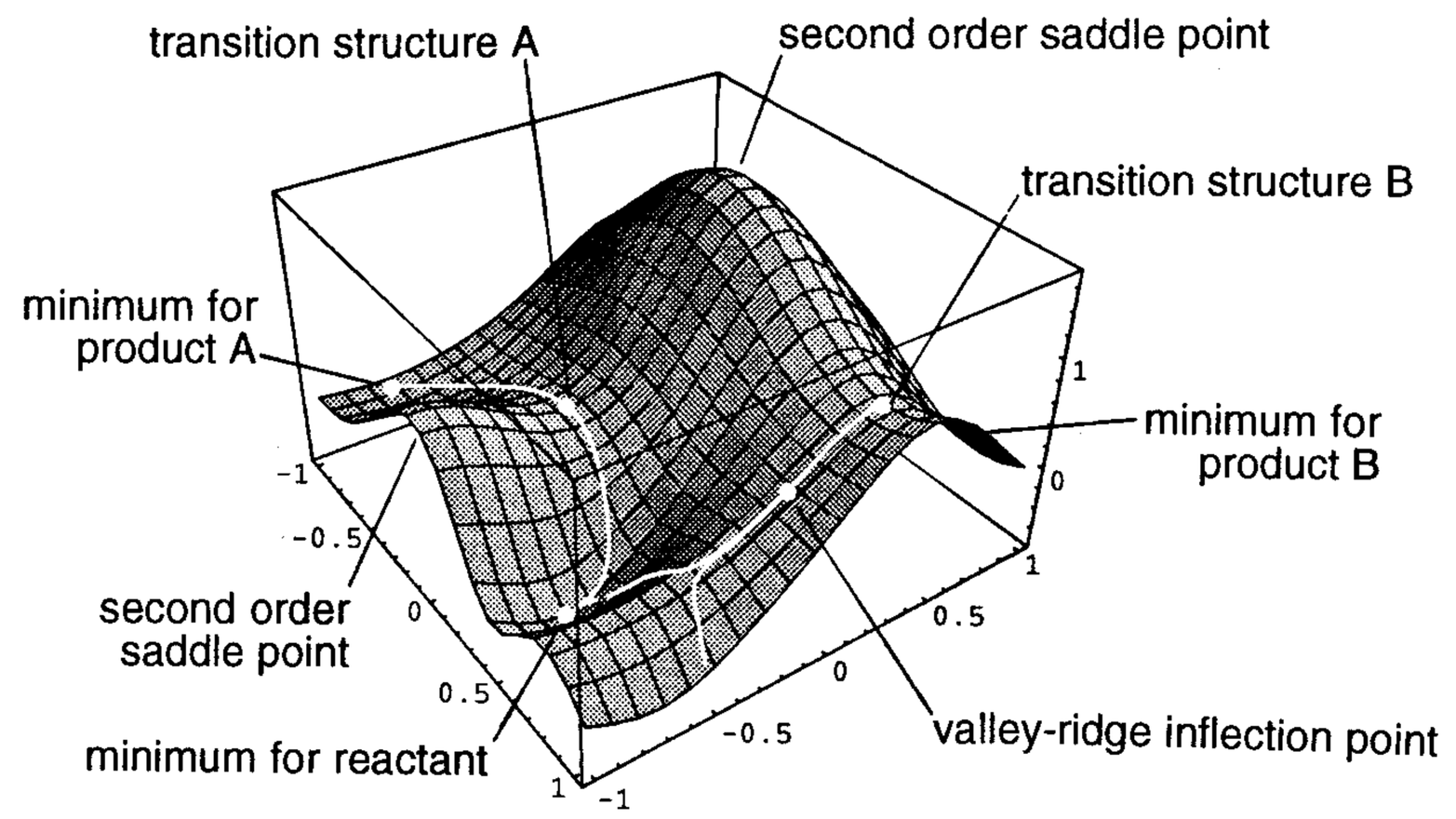


Fig. 1. A model potential energy surface[25] illustrating minima, transition structures, second order saddle points, a valley–ridge inflection point and reaction paths.

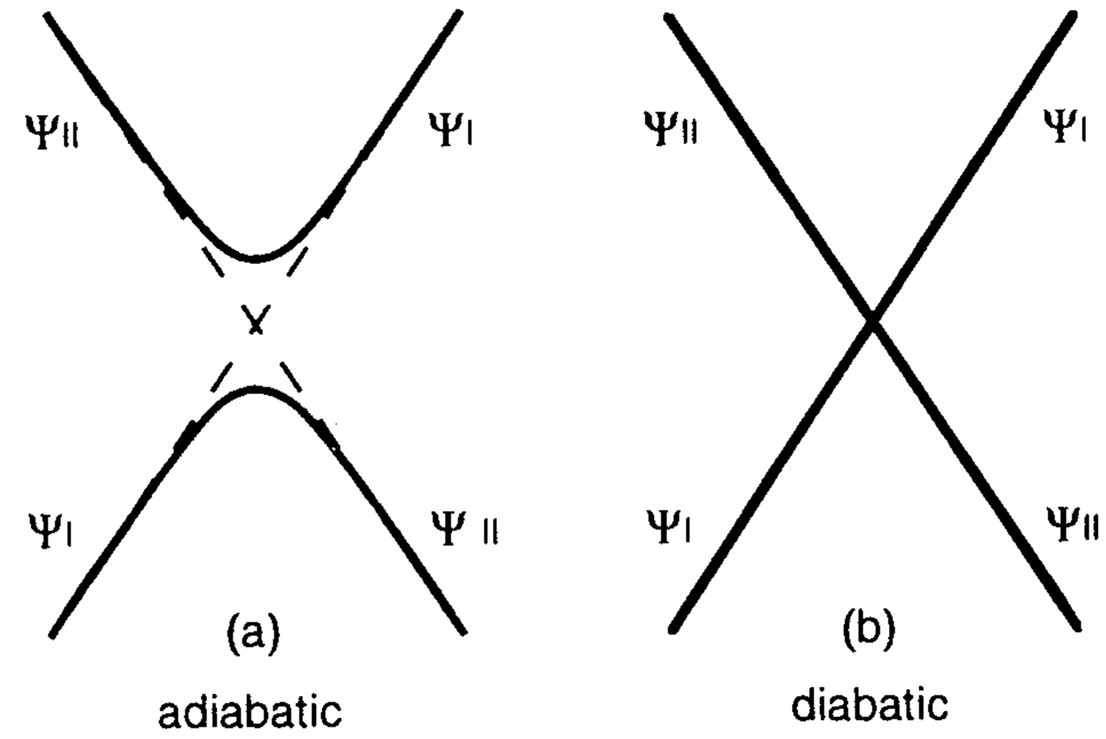
A *transition structure* is the highest point along the lowest energy reaction path connecting reactants and products. Thus, a transition structure must be a maximum in one direction (along the reaction path). A transition structure must also be a minimum in all other directions, since it is the lowest energy reaction path, i.e., any displacement perpendicular to the reaction path at the transition structure must lead to a higher energy. A stationary point that is a maximum in one direction and a minimum in all orthogonal directions is a *first-order saddle point.* As with minima, to test whether a point is a saddle point, one must first diagonalize the Hessian. If there is one and only one negative eigenvalue and all the remaining eigenvalues are positive, then a first-order saddle point has been found. Equivalently, a first-order saddle point has one and only one imaginary frequency and all other vibrational frequencies are real. The eigenvector or normal mode of vibration corresponding to the negative eigenvalue or the imaginary frequency is called the *transition vector* and points along the reaction path between the reactants and products. There can be more

than one saddle point along a reaction path if there are intermediates. If there is only one reaction path connecting reactants and products, then the transition structure is the highest point along the path. If there are multiple paths between the same reactants and products, then the transition structure is the maximum of the lowest energy reaction path between reactants and products. Figure 1 shows two different saddle points, each of which is a transition structure in the region of the surface shown in the figure. A transition structure refers to the geometric structure of the first-order saddle point; a transition state includes the vibrational and rotational states of the saddle point structure. However, these terms are often used interchangeably in the literature.

A *second-order saddle point* is characterized by two negative eigenvalues of the Hessian or two imaginary frequencies. As shown in Fig. 1, this corresponds to a local maximum in two of the coordinates of the potential energy surface. If a reaction path were to go through a second-order saddle point, then a lower energy pathway could always be found by displacing the path away from the maximum and going around the local hill top. Note, however, that for higher dimensional potential energy surfaces, descending along a ridge from a second-order saddle point can lead to another second-order saddle point rather than to a first-order saddle point.

2.3. *Seams, Conical Intersections and Avoided Crossings*

In some regions, ground state and excited state potential surfaces may come close together, touch or cross, and molecules can move readily from one surface to the other.[19–22] In one dimension, the Born–Oppenheimer treatment yields *adiabatic* energy curves which do not cross if they are of the same symmetry; part (a) of Scheme 1 shows an avoided crossing of this type. When potential energy surfaces come close, the nature of the electronic wave function can change rapidly and the Born–Oppenheimer approximation may no longer be valid. In Scheme 1, the lower surface is dominated by electronic configuration Ψ_{I} before the avoided crossing and by Ψ_{II} after. An alternate representation involves *diabatic* surfaces that can cross, shown in part (b) of Scheme 1. Diabatic surfaces preserve the dominant characteristics of the electronic wave functions, but the surfaces have non-zero interaction matrix elements which control the rate of transition from one surface to the other. The treatment of rate process involving intersections or weakly avoided crossings must take into account these



Scheme 1

non-Born–Oppenheimer or non-adiabatic effects. Non-adiabatic rate processes are discussed in detail elsewhere in this volume.[19] These treatments require the location of the crossing or closest approach between the curves and need the shape of the curves in this vicinity.

Two-dimensional potential energy surfaces of different spatial or spin symmetry can intersect in a one-dimensional line or *seam*, as shown in Fig. 2(a). For n-dimensional surfaces, the seam is $(n-1)$-dimensional. The lowest point on the seam is needed for computing rate processes and must be found by a geometry optimization constrained to follow the seam. If a pair of two-dimensional diabatic potential energy surfaces can interact, they can form a *conical intersection*, as shown in Fig. 2(b). At the apex of a conical intersection, the surfaces cross because the interaction matrix element is zero. Any displacement from this point leads to a non-zero separation between the diabatic surfaces and/or a non-zero interaction matrix element between the surfaces, resulting in an avoided crossing between the adiabatic surfaces. For n-dimensional surfaces, the conical intersection is $(n-2)$-dimensional. A third type of surface interaction is called a *funnel* and is shown in Fig. 2(c). Two surfaces approach each other closely but do not touch. An unconstrained geometry optimization is needed to locate the point of closest approach so that transition probabilities from one surface to the other can be computed. A conical intersection can be considered a special case of a funnel.

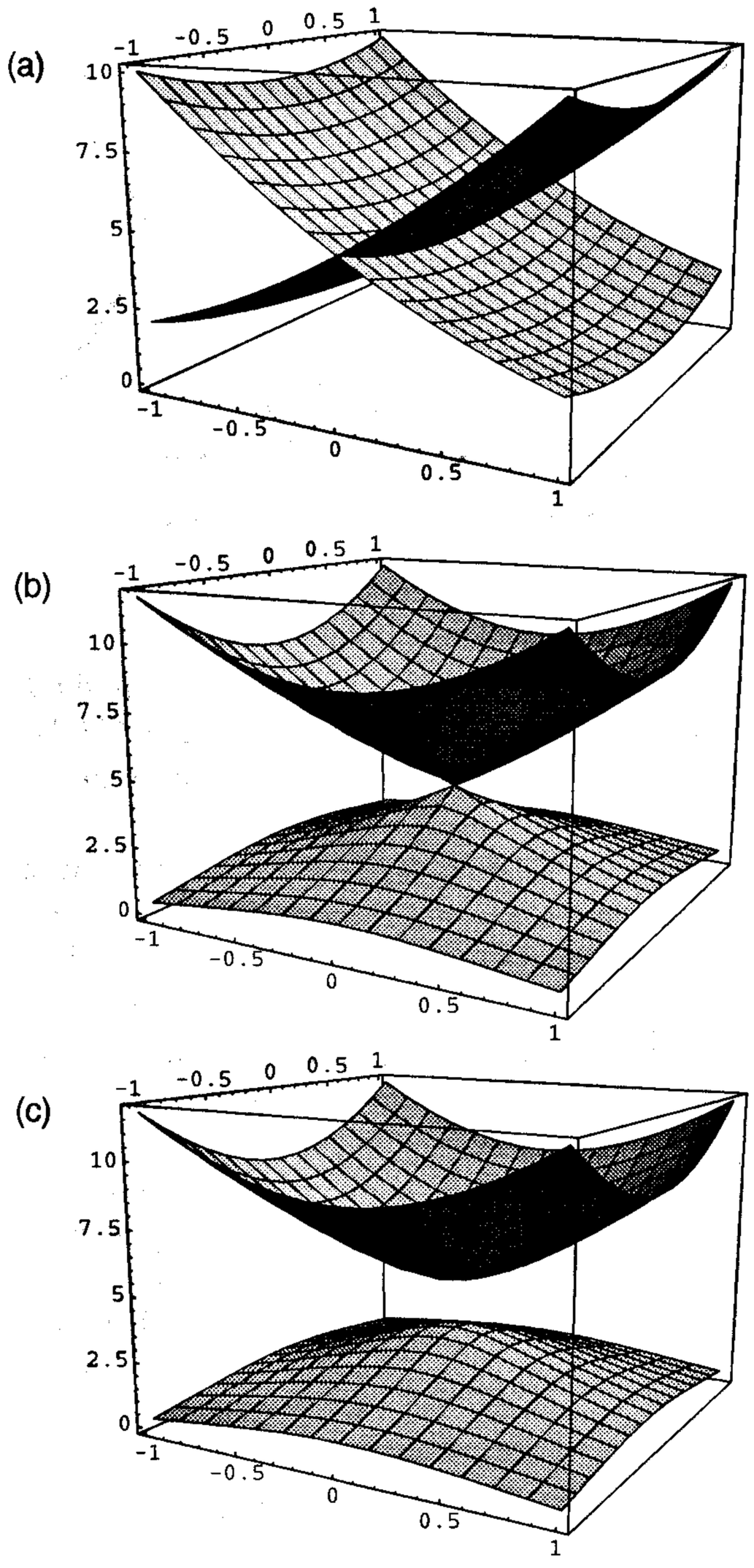


Fig. 2. Interaction between two model potential energy surfaces[25] showing (a) a seam, (b) a conical intersection and (c) a weakly avoided conical intersection or funnel.

2.4. *Reaction Paths*

A *reaction path* can be defined as the steepest descent path from a transition structure down to reactants and down to products.[26] If $\mathbf{x}(s)$ is the path and s is the arc length along the path, then $\mathbf{v}^0$, the tangent to reaction path, is given by

$$\mathbf{v}^0 = \frac{\partial \mathbf{x}(s)}{\partial s} = -\frac{\mathbf{g}(s)}{|\mathbf{g}(s)|}, \tag{1}$$

where $\mathbf{g}(s)$ is the gradient at $\mathbf{x}(s)$. The path can be followed by integrating Eq. (1) using any of a variety of numerical methods. Since algorithms for following reaction paths bear a strong resemblance to geometry optimization, they are included in this chapter. The reaction path is defined only in a downhill fashion; it cannot be followed uphill unambiguously, since every direction from a minimum is uphill, and only a very limited number of paths lead to first-order saddle points. A reaction path descending from one saddle point may merge with a path from another saddle point. A valley descending from a transition state or ascending to a transition state can also *bifurcate*, i.e., branch into two separate valleys[27] (see Fig. 1).

A reaction path depends on the coordinate system, since the gradient and displacement transform in different ways (covariant and contravariant, respectively). One system, mass-weighted Cartesian coordinates, has unique properties that makes it desirable from a dynamical point of view. In mass-weighted Cartesians, the instantaneous acceleration of a point on the path is in the same direction as the displacement of the point downward along the path. The reaction path in mass-weighted Cartesian coordinates is termed the *intrinsic reaction coordinate* (IRC).[26]

Some methods for calculating reaction rates, such as variational transition state theory and reaction path Hamiltonians,[28] require the curvature of the reaction path as well as the path itself. The curvature, $\mathbf{v}^1$, is given by

$$\mathbf{v}^1 = \frac{\partial^2 \mathbf{x}(s)}{\partial s^2} = -\frac{(\mathbf{H}\mathbf{v}^0 - (\mathbf{v}^{0t}\mathbf{H}\mathbf{v}^0)\mathbf{v}^0)}{|\mathbf{g}(s)|}, \tag{2}$$

where $\mathbf{H}$ is the Hessian. Projected vibrational frequencies and coupling matrix elements between motion along the path and vibrations perpendicular to the path can be calculated easily from the tangent and curvature vectors.

Although the reaction path in any coordinate system is uniquely defined, it is not possible to tell whether an arbitrary point on the surface is on a

reaction path without first finding the saddle point and following the path downhill. A *gradient extremal path*[29] is a different type of path that is locally defined and connects minima and saddle points. On a gradient extremal, the gradient is an eigenvector of the Hessian:

$$\mathbf{H}(s)\mathbf{g}(s) = c\mathbf{g}(s)\,. \tag{3}$$

Because this definition does not require knowledge of the transition structure, any point on the potential energy surface can be displaced toward the closest gradient extremal path and then the gradient extremal can be followed uphill toward the transition state. The two drawbacks of gradient extremals are that they require the Hessian, and that they are a less direct route to the saddle point than steepest descent paths since they have a tendency to wander about the potential energy surface.[29]

3. Algorithms for Finding Minima

The geometry optimization task encountered in molecular orbital calculation translates into an unconstrained minimization of a continuous function, a problem discussed at length in numerical analysis texts.[5] The potential energy surface in a geometry optimization problem is called the *objective function* in numerical analysis, the energy at the equilibrium geometry is the *minimum* of the objective function and the equilibrium geometry is called the *minimizer* of the objective function. Since constrained and unconstrained minimizations are at the heart of all the methods for locating features on potential energy surfaces considered in this chapter, the various aspects of minimization algorithms are discussed in somewhat greater detail. Standard texts in numerical analysis[5] and previous reviews of geometry optimization[12] should be consulted if a more extensive treatment is needed. Optimization methods can be divided into three groups, depending on whether the only objective function is used, or the function and its gradient, or the function, gradient and Hessian.

Energy-only methods are the most widely applicable, but are the slowest to converge. The axial iteration and univariate search methods optimize one coordinate at a time; if the coordinates are strongly coupled, many iterations through all of the coordinates may be needed to achieve satisfactory convergence. The simplex method is somewhat better at handling coupled coordinates, but the number of energy calculations required for convergence

in an n-dimensional case is still proportional to n^2. For optimizations in more than three or four dimensions when no analytical gradients are available, it is often better to calculate the gradients numerically and to use one of the gradient-based optimization algorithms.

If gradients are available, the rate of convergence of an optimization can be greatly improved, since each step provides $n+1$ additional pieces of information about the potential energy surface (i.e., the energy and n components of the gradient at the current point). The steepest descent method is the simplest approach — at each step, it finds a minimum in the downhill direction by searching along the negative of the gradient vector. It is effective at lowering the energy quickly, but is slow to converge to the final minimum. Conjugate gradient and quasi-Newton methods are much more efficient at finding minima. They are based on a local quadratic model of the objective function, and search for the minimum using conjugate steps rather than steepest descent steps. It is simple to prove that these methods converge to a minimum in $n+1$ or fewer steps for an n-dimensional quadratic function.[5]

The flow chart of a typical quasi-Newton-based geometry optimization is shown in Fig. 3. The calculation starts with a choice of coordinate system, and an initial estimate of the geometry and the Hessian (these practical considerations are discussed in Sec. 7 and Ref. 18). Each cycle in the optimization consists of calculating the energy and gradient, minimizing along the search direction, updating the current estimate of the Hessian, determining a new search direction and checking for convergence. Optimization methods differ by the accuracy of the line search, the updating scheme used for the Hessian, the determination of the new search direction and next estimate of the minimizer. These aspects of the optimization problem are discussed below.

Hessian-based methods (Newton, Gill–Murray, etc.) are useful when convergence difficulties are encountered by other approaches. If the Hessian cannot be calculated analytically, it can be obtained by numerically differentiating the gradients. Determination of the search direction and control of the step size are the same as for gradient-based methods. Thus, Hessian-based methods can be treated as a special case of quasi-Newton algorithms, with the Hessian updating step replaced by an explicit calculation of the Hessian. Often, the desired rate of convergence for difficult optimizations can be achieved by recalculating the Hessian every few

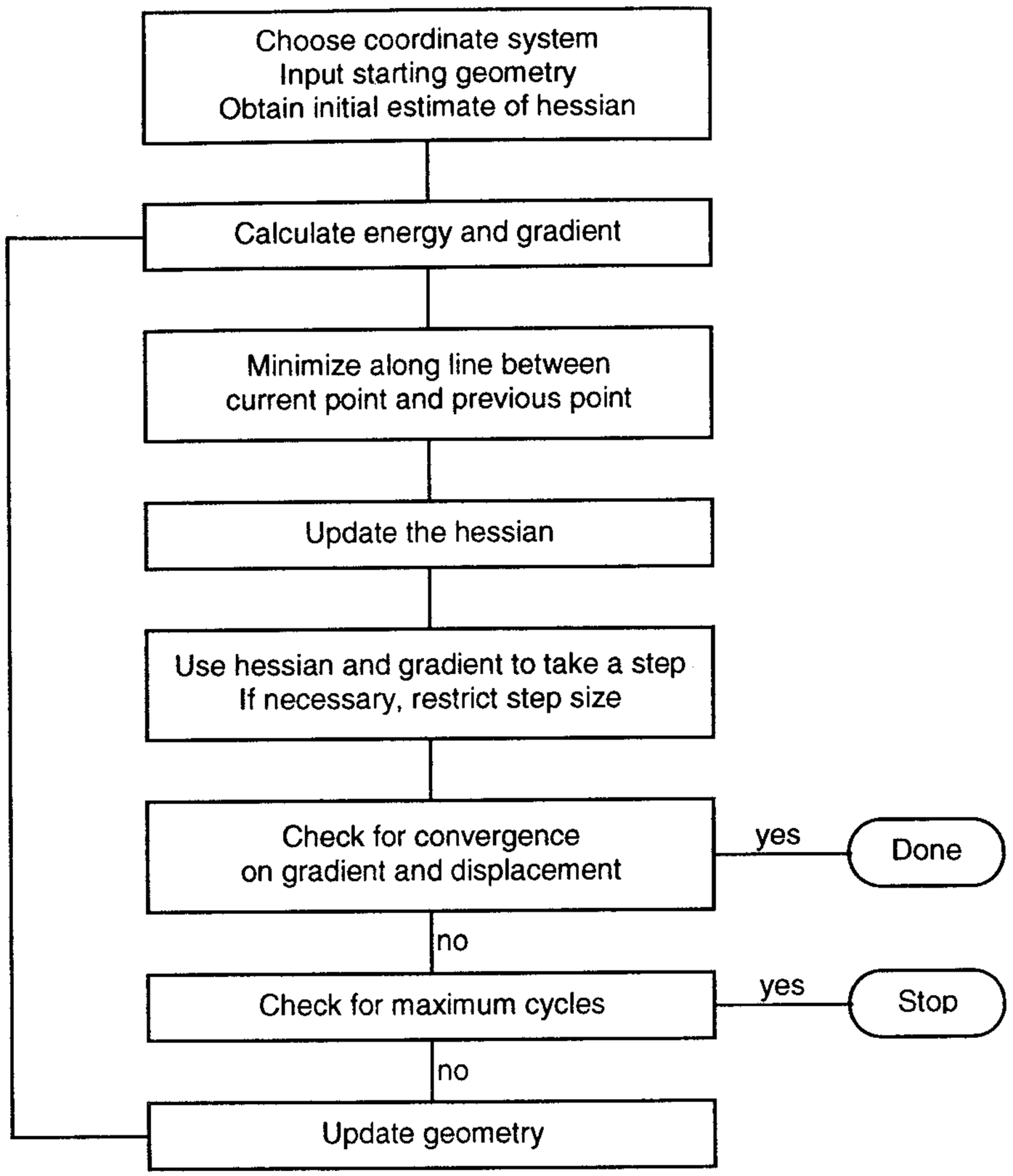


Fig. 3. Flow chart for quasi-Newton algorithms for geometry optimization.

steps rather than at each step, or perhaps by calculating the Hessian only at the beginning.

Simulated annealing methods[30] are considered to be outside the scope of this chapter. In these approaches, one starts with an ensemble of systems and the atoms are given an appropriate amount of kinetic energy; each system in the ensemble is allowed to propagate by Monte Carlo or molecular dynamics methods. The kinetic energy is drained from the systems at a prescribed cooling rate and they settle into the deepest minima on the surface.

Simulated annealing is useful for finding global minima in high-dimensional systems, but such a search can require 10^4 to 10^8 steps. It has been applied successfully to molecular mechanics treatments of macromolecules, but is not practical for molecular orbital calculations at this time.

3.1. *Hessian Update*

A quasi-Newton optimization begins with an approximate Hessian, $\mathbf{B}$ (to be distinguished from $\mathbf{H}$, the actual Hessian), and uses a quadratic model of the potential energy surface to predict the next estimate of the minimizer:

$$E(\mathbf{x}) = E(\mathbf{x}_i) + \mathbf{g}_i^t(\mathbf{x} - \mathbf{x}_i) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_i)^t\mathbf{B}_i(\mathbf{x} - \mathbf{x}_i)\,, \tag{4}$$

$$\mathbf{g}_{i+1} = \mathbf{g}(\mathbf{x}_{i+1}) = \mathbf{g}_i + \mathbf{B}_i(\mathbf{x}_{i+1} - \mathbf{x}_i) = 0\,; \quad \mathbf{x}_{i+1} = \mathbf{x}_i - \mathbf{B}_i^{-1}\mathbf{g}_i\,. \tag{5}$$

The quadratic model must be improved during the course of the optimization by updating the estimated Hessian, $\mathbf{B}$, or its inverse, $\mathbf{B}^{-1}$. The number of degrees of freedom in a geometry optimization by *ab initio* molecular orbital methods can be up to a few hundred. Nevertheless, the cost of storing, updating and inverting the Hessian is insignificant, since molecular orbital calculations can require 10^6 to 10^{12} floating point operations per energy plus gradient evaluation, and 10^6 to 10^9 words of storage.

At any step, the approximate Hessian (or its inverse) must be such that the quadratic model fits the gradient at the current and previous points. This requirement reduces to:

$$\mathbf{B}_i\Delta\mathbf{x}_i = \Delta\mathbf{g}_i \quad \text{or} \quad \mathbf{B}_i^{-1}\Delta\mathbf{g}_i = \Delta\mathbf{x}_i\,, \tag{6}$$

where $\Delta\mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_{i-1}$ and $\Delta\mathbf{g}_i = \mathbf{g}_i - \mathbf{g}_{i-1}$, and is sometimes called the quasi-Newton condition. For quadratic surfaces, many of the update methods discussed below also satisfy the quasi-Newton condition for previous steps:

$$\mathbf{B}_i\Delta\mathbf{x}_k = \Delta\mathbf{g}_k \quad \text{or} \quad \mathbf{B}_i^{-1}\Delta\mathbf{g}_k = \Delta\mathbf{x}_k,\ k < i\,. \tag{7}$$

This is known as the heredity property. A further requirement is that the Hessian remains symmetric. Updating formulae can be classified according to the rank of the correction to $\mathbf{B}$ or $\mathbf{B}^{-1}$.

The rank one updating formula can be written as:

$$\mathbf{B}_i = \mathbf{B}_{i-1} + \frac{(\Delta\mathbf{g}_i - \mathbf{B}_{i-1}\Delta\mathbf{x}_i)(\Delta\mathbf{g}_i - \mathbf{B}_{i-1}\Delta\mathbf{x}_i)^t}{(\Delta\mathbf{g}_i - \mathbf{B}_{i-1}\Delta\mathbf{x}_i)^t\Delta\mathbf{x}_i}\,, \tag{8}$$

$$\mathbf{B}_i^{-1} = \mathbf{B}_{i-1}^{-1} + \frac{(\Delta\mathbf{x}_i - \mathbf{B}_{i-1}^{-1}\Delta\mathbf{g}_i)(\Delta\mathbf{x}_i - \mathbf{B}_{i-1}^{-1}\Delta\mathbf{g}_i)^t}{(\Delta\mathbf{x}_i - \mathbf{B}_{i-1}^{-1}\Delta\mathbf{g}_i)^t\Delta\mathbf{g}_i}, \tag{9}$$

and is associated with names such as Broyden,[31] Murtagh and Sargent[32] and others. These updating formulas suffer two disadvantages: (a) they do not necessarily maintain positive definiteness of the Hessian, which is needed to ensure that each step is successively closer to the minimum, and (b) the denominator may become very small or zero, causing numerical difficulties. Although usually not written in this fashion, the conjugate gradient methods[33,34] (Eqs. (20) and (21)) also correspond to rank one formulae, e.g.

$$\mathbf{B}_i^{-1} = \mathbf{B}_i^{-1} + \frac{(\mathbf{g}_i - \beta\mathbf{s}_{i-1} + \mathbf{B}_{i-1}^{-1}\mathbf{g}_i)(\mathbf{g}_i - \beta\mathbf{s}_{i-1} + \mathbf{B}_{i-1}^{-1}\mathbf{g}_i)^t}{(\mathbf{g}_i - \beta\mathbf{s}_{i-1} + \mathbf{B}_{i-1}^{-1}\mathbf{g}_i)^+\mathbf{g}_i}; \quad \mathbf{B}_0 = \mathbf{I} \tag{10}$$

Conjugate gradient methods normally do not store the Hessian, but compute the search direction directly from the current and previous gradients (see below). Because they require storage for only a displacement vector and one or two gradient vectors, conjugate gradients are the methods of choice for very large optimization problems, where storage of the inverse Hessian is not possible.

The symmetric, rank two updating formula of Powell[35] is:

$$\mathbf{B}_i = \mathbf{B}_{i-1} + \frac{(\Delta\mathbf{g}_i - \mathbf{B}_{i-1}\Delta\mathbf{x}_i)\Delta\mathbf{x}_i^t + \Delta\mathbf{x}_i(\Delta\mathbf{g}_i - \mathbf{B}_{i-1}\Delta\mathbf{x}_i)^t}{\Delta\mathbf{x}_i^t\Delta\mathbf{x}_i} - \frac{(\Delta\mathbf{x}_i^t\Delta\mathbf{g}_i - \Delta\mathbf{x}_i^t\mathbf{B}_{i-1}\Delta\mathbf{x}_i)\Delta\mathbf{x}_i\Delta\mathbf{x}_i^t}{(\Delta\mathbf{x}_i^t\Delta\mathbf{x}_i)^2}. \tag{11}$$

This update is usually not used for minimization because it is not guaranteed to give a positive definite update for a positive definite Hessian. However, this property makes it quite useful for transition structure optimization (see Sec. 4).

The DFP and BFGS formulas are the most frequently used rank two updating schemes; both are members of the Broyden family of updating formulas. The Davidon–Fletcher–Powell update[36] for the inverse Hessian is:

$$\mathbf{B}_i^{-1} = \mathbf{B}_{i-1}^{-1} + \frac{\Delta\mathbf{x}_i\Delta\mathbf{x}_i^t}{\Delta\mathbf{x}_i^t\Delta\mathbf{g}_i} - \frac{\mathbf{B}_{i-1}^{-1}\Delta\mathbf{g}_i\Delta\mathbf{g}_i^t\mathbf{B}_{i-1}^{-1}}{\Delta\mathbf{g}_i^t\mathbf{B}_{i-1}^{-1}\Delta\mathbf{g}_i}. \tag{12}$$

The Broyden–Fletcher–Goldfarb–Shanno formula[37] for the inverse Hessian is:

$$\mathbf{B}_i^{-1} = \mathbf{B}_{i-1}^{-1} + \frac{\Delta\mathbf{x}_i\Delta\mathbf{x}_i^t}{\Delta\mathbf{x}_i^t\Delta\mathbf{g}_i} - \frac{\mathbf{B}_{i-1}^{-1}\Delta\mathbf{g}_i\Delta\mathbf{g}_i^t\mathbf{B}_{i-1}^{-1}}{\Delta\mathbf{g}_i^t\mathbf{B}_{i-1}^{-1}\Delta\mathbf{g}_i} + (\Delta\mathbf{g}_i^t\mathbf{B}_{i-1}^{-1}\Delta\mathbf{g}_i)\mathbf{w}\mathbf{w}^t, \tag{13}$$

$$\mathbf{w} = \frac{\Delta\mathbf{x}_i}{\Delta\mathbf{x}_i^t\Delta\mathbf{g}_i} - \frac{\mathbf{B}_{i-1}^{-1}\Delta\mathbf{g}_i}{\Delta\mathbf{g}_i^t\mathbf{B}_{i-1}^{-1}\Delta\mathbf{g}_i}.$$

These formulae are dual or complementary, in that the BFGS update for the Hessian (Eq. (14)) resembles the DFP update for the inverse Hessian (Eq. (12)):

$$\mathbf{B}_i = \mathbf{B}_{i-1} + \frac{\Delta\mathbf{g}_i\Delta\mathbf{g}_i^t}{\Delta\mathbf{g}_i^t\Delta\mathbf{x}_i} - \frac{\mathbf{B}_{i-1}\Delta\mathbf{x}_i\Delta\mathbf{x}_i^t\mathbf{B}_{i-1}}{\Delta\mathbf{x}_i^t\mathbf{B}_{i-1}\Delta\mathbf{x}_i}. \tag{14}$$

Likewise, the DFP formula for the Hessian update can be obtained by interchanging $\Delta\mathbf{x}$ and $\Delta\mathbf{g}$, $\mathbf{B}$ and $\mathbf{B}^{-1}$ in the BFGS formula for updating the inverse. The BFGS formulae, Eqs. (13) and (14), are generally acknowledged to be superior to DFP.[5]

Another updating scheme that has found widespread use within the Gaussian series of programs is a method devised by the author a number of years ago.[38] The gradients in the space spanned by the current and all previous displacements are used to computed the Hessian in this space:

$$b_{jk} = \left[(\mathbf{g}_j - \mathbf{g}_i)^t\mathbf{r}_k - \sum_{m=j+1}^{i-1} b_{mk}((\mathbf{x}_j - \mathbf{x}_i)^t\mathbf{r}_m)\right] \Big/ (\mathbf{x}_j - \mathbf{x}_i)^t\mathbf{r}_j;$$

$$b_{kj} = b_{jk} \quad j \le k = i-1,\ i-2, \ldots, \tag{15}$$

$$\mathbf{r}'_j = (\mathbf{x}_j - \mathbf{x}_i) - \sum_{m=j+1}^{i-1} \mathbf{r}_m((\mathbf{x}_j - \mathbf{x}_i)^t\mathbf{r}_m); \quad \mathbf{r}_j = \mathbf{r}'_j/|\mathbf{r}'_j| \quad j = i-1, i-2, \ldots, \tag{16}$$

$$\mathbf{B}_i = \mathbf{B}_{i-1} + \sum_{jk}(b_{jk} - \mathbf{r}_j\mathbf{B}_{i-1}\mathbf{r}_k)\mathbf{r}_j\mathbf{r}_k^t. \tag{17}$$

The vectors $\mathbf{r}$ that span the small space are obtained by Schmidt orthogonalization, starting from the displacement between the current point and

previous point. The upper triangle of the Hessian is computed in the small space and is symmetrized. The difference between the new Hessian and the old Hessian in the small space is transformed back to the full space of the optimization variables and used to update the Hessian. At step i, this results in a rank i update to the Hessian.

3.2. *Search Direction and Step Size Control*

Finding the next search direction and determining how far to step in that direction toward the minimum are intimately connected, and are therefore discussed together. As mentioned above, the simplest and least efficient of the gradient approaches is the steepest descent method. The downhill direction is along the negative of the gradient; α is chosen to minimize the energy along this line.

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha \mathbf{g}_i \,. \tag{18}$$

However, after minimizing in the current direction, the energy is no longer a minimum in the directions searched previously.

If the current point, $\mathbf{x}_i$, is a minimum with respect to the previous search directions, then $\Delta \mathbf{x}_j^t \mathbf{g}_i = 0$. Since the estimated gradient in the new search direction is $\mathbf{g}_i + \mathbf{B}\Delta x_{i+1}$, one must require $\Delta \mathbf{x}_j^t \, (\mathbf{g}_i + \mathbf{B}\Delta \mathbf{x}_{i+1}) = \Delta \mathbf{x}_j^t \mathbf{B} \Delta \mathbf{x}_{i+1} = 0$ to retain the minima in previous directions, i.e., the new search direction must be conjugate to the previous search directions. Any method that generates conjugate directions and uses an accurate line search can be proven to converge in $n+1$ or fewer steps for an n-dimensional quadratic function. The conjugate gradient methods are the simplest of this class:

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha \mathbf{s}_i \,, \tag{19}$$

$$\mathbf{s}_i = -\mathbf{g}_i + \frac{\mathbf{g}_i^t \mathbf{g}_i}{\mathbf{g}_{i-1}^t \mathbf{g}_{i-1}} \mathbf{s}_{i-1} \,, \quad \text{(Fletcher–Reeves}^{33}\text{)} \tag{20}$$

$$\mathbf{s}_i = -\mathbf{g}_i + \frac{(\mathbf{g}_i - \mathbf{g}_{i-1})^t \mathbf{g}_i}{\mathbf{g}_{i-1}^t \mathbf{g}_{i-1}} \mathbf{s}_{i-1} \,. \quad \text{(Polak–Ribiere}^{34}\text{)} \tag{21}$$

The Polak–Ribiere formula with an accurate line search ($|\mathbf{g}_{i+1}^t \Delta \mathbf{x}_{i+1}| \leq 0.1 \mathbf{g}_i^t \Delta \mathbf{x}_{i+1}$) appears to be better for a number of functions. Since the Hessian is not used and only a few vectors are stored, conjugate gradient algorithms are the methods of choice for very large optimization problems where storage space is at a premium.

Quasi-Newton methods use an updated Hessian to determine the search direction (Newton or Newton–Raphson step):

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha \mathbf{B}^{-1} \mathbf{g}_i \,. \tag{22}$$

The Hessian must be positive definite to yield a downhill search direction. If the Hessian is updated by one of the methods discussed above, the search directions are conjugate and the optimization converges in $n+1$ iterations for an n-dimensional quadratic function. For non-quadratic functions, quasi-Newton methods converge more rapidly than conjugate gradient methods.

An alternative method for obtaining the next estimate of the optimum geometry is used by Pulay.[39] A linear combination of the current and previous points is chosen so that the predicted displacement from this point is smallest in a least squares sense:

$$\mathbf{x}_i' = \sum_{j \leq i} c_j \mathbf{x}_j \text{ such that } \sum_{j \leq i} c_j = 1 \text{ and } \left(\mathbf{B}^{-1} \sum_{j \leq i} c_j \mathbf{g}_j \right)^2 \text{ is a minimum} \,, \tag{23}$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i' - \mathbf{B}^{-1} \sum_{j \leq i} c_j \mathbf{g}_j \,. \tag{24}$$

Normally, no additional line search is needed and the approximate Hessian can be used with or without updating. The performance of this method is equivalent to quasi-Newton methods for a variety of geometry optimization problems.[39,40]

Another approach has been proposed by McKelvey and Hamilton.[41] Instead of searching along a line, they suggest searching along the steepest descent path:

$$\frac{d\mathbf{x}(t)}{dt} = -\mathbf{g}_i - \mathbf{B}(\mathbf{x}(t) - \mathbf{x}_i) \,, \quad \text{and} \quad \mathbf{x}_{i+1} \text{ is the minimum along } \mathbf{x}(t) \,. \tag{25}$$

At small displacements, this gives a steepest descent step; the limit as $t \to \infty$ is a Newton step (Eq. (22), $\alpha = 1$).

Many minimization algorithms require a line search in the descent direction. If the energy calculation is much less expensive than the gradient

evaluation (as in the case for numerical gradients), then it may be worthwhile to do an accurate line search in order to minimize the number of gradient calculations and Newton steps. The more typical situation in geometry optimization by *ab initio* molecular orbital methods is that the cost of gradients is about the same as for the energy.[1–4] In this case, it is more efficient, in terms of the most information about potential energy surface in the least amount of computer time, to calculate and use the gradient each time the energy is evaluated. This implies that the best strategy is to use an approximate line search and to take a Newton step whenever possible.

The minimization in the search direction starts with a displacement in that direction. For quasi-Newton methods, the first displacement in the line search is chosen with $\alpha = 1$ in Eq. (22). Additional points are calculated until the search is terminated, usually based on a reduction in the gradient in the search direction. If the convergence criterion is based on the gradient along the search direction, then $|\mathbf{g}_{i+1}^t \Delta\mathbf{x}_{i+1}| \leq \sigma |\mathbf{g}_i^t \Delta\mathbf{x}_{i+1}|$, and $\sigma = 0.1$ corresponds to an accurate search, whereas $\sigma = 0.9$ is appropriate for an approximate line search. An approximate line search is sufficient for most quasi-Newton methods and is usually achieved on the first step (i.e. $\alpha = 1$). A more accurate location of the minimum in the line search can be obtained by polynomial interpolation. A quartic polynomial fitted to the energy and gradients at $\alpha = 0$ and $\alpha = 1$ and constrained to have only one minimum has been found to be quite effective.[38]

During the early phases of an optimization, the Newton step can lead to a displacement that is larger than the perceived quadratic region of the potential energy surface. Rather than risk a large, potentially very bad step, it is usually better to take a smaller, more controlled step. The simplest approach is to reduce the step by multiplying it by a scale factor. Alternatively, one could adjust only the largest components of the displacement, either in terms of the optimization variables or in terms of components along the eigenvectors of the Hessian. A better approach is to introduce a trust radius[5] τ. Any step larger than the trust radius is reduced to the trust radius and the direction is adjusted to give the lowest energy. This amounts to a constrained minimization with a fixed length displacement and can be computed using Lagrange multipliers:

$$\text{minimize } E = E_i + \mathbf{g}_i^t \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^t \mathbf{B} \Delta\mathbf{x}, \text{ subject to } \Delta\mathbf{x}^t \Delta\mathbf{x} = \tau^2 , \tag{26}$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \sum_{j=1}^{n} \frac{\mathbf{v}_j (\mathbf{v}_j^t \mathbf{g}_i)}{(b_i - \lambda)}, \text{ where } \lambda < b_1 \text{ and satisfies } \sum_{j=1}^{n} \frac{(\mathbf{v}_j^t \mathbf{g}_i)^2}{(b_i - \lambda)^2} = \tau^2 , \tag{27}$$

where b_j are the eigenvalues of $\mathbf{B}$ (b_1 is the lowest) and $\mathbf{v}_j$ are the corresponding eigenvectors. The trust radius should be decreased if the energy difference predicted by the quadratic step is much different than the actual energy difference (i.e., the step is outside the quadratic region of the function), or can be increased if the two are comparable (i.e., the step is inside the quadratic region):

$$r = \frac{E_{i+1} - E_i}{\mathbf{g}_i \Delta\mathbf{x}_{i+1} + \frac{1}{2}\Delta\mathbf{x}_{i+1}^t \mathbf{B} \Delta\mathbf{x}_{i+1}};$$

$$\tau = \tau/\alpha \text{ if } r < r_1 \text{ or } r > 2 - r_1 ; \quad \tau = \alpha\tau \text{ if } r_2 < r < 2 - r_2 . \tag{28}$$

For a conservative trust radius update, one might choose $r_1 = 0.7$, $r_2 = 0.85$, $\alpha = 1.5$.

Another approach to controlling the step size in an optimization involves replacing the quadratic model inherent to quasi-Newton methods with a rational function (hence the name rational function optimization or RFO).[42]

$$E - E_i = \frac{\mathbf{g}_i^t \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^t \mathbf{B} \Delta\mathbf{x}}{1 + \Delta\mathbf{x}^t \mathbf{S} \Delta\mathbf{x}} = \frac{1}{2} \frac{[\Delta\mathbf{x}^t \quad 1] \begin{bmatrix} \mathbf{B} & \mathbf{g}_i \\ \mathbf{g}_i^t & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ 1 \end{bmatrix}}{[\Delta\mathbf{x}^t \quad 1] \begin{bmatrix} \mathbf{S} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ 1 \end{bmatrix}} \tag{29}$$

$$\frac{dE}{d\Delta\mathbf{x}} = \begin{bmatrix} \mathbf{B} & \mathbf{g}_i \\ \mathbf{g}_i^t & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ 1 \end{bmatrix} - 2(E - E_i) \begin{bmatrix} \mathbf{S} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ 1 \end{bmatrix} = 0 \tag{30}$$

$$\begin{bmatrix} \mathbf{B} & \mathbf{g}_i \\ \mathbf{g}_i^t & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{S} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ 1 \end{bmatrix} ; \quad \lambda = 2(E - E_i)$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i - (\mathbf{B} - \lambda\mathbf{S})^{-1} \mathbf{g}_i \tag{31}$$

The choice $\mathbf{S} = \alpha^2 \mathbf{I}$ is equivalent to scaling the displacement by α,[42] i.e., $\Delta\mathbf{x}'$ computed with $\mathbf{S} = \alpha^2 \mathbf{I}$ is equal to $\alpha\Delta\mathbf{x}$ where $\Delta\mathbf{x}$ is computed with $\mathbf{S} = \mathbf{I}$. If the constant α is chosen so that $\Delta\mathbf{x}^t \Delta\mathbf{x} = \tau^2$, then this method of step size control is the same as the trust radius method.

Under some circumstances, it may be necessary to constrain parts of a molecule during a geometry optimization. If non-redundant internal coordinates are used for the optimization, then any of the bond lengths, angles or torsions used in the definition of the molecular geometry can be constrained to their initial values by excluding them from the optimization space. When the constraints involve linear combinations of internal coordinates, or when Cartesian coordinates or redundant internal coordinates are used for the optimization, the constraints cannot be applied simply by fixing the values of a few coordinates. Instead, the constraints must be imposed by Lagrangian multipliers,[43a] projector methods[43b] or a penalty function approach.[43c]

4. Algorithms for Finding Transition States

Finding transition states for chemical reactions is at least an order of magnitude more difficult than finding equilibrium geometries or minima. Quasi-Newton methods are very robust and efficient for minimization, but when such methods are modified for locating saddle points, they have a much smaller radius of convergence. For minima, empirical rules for molecular structures and chemical intuition about bonding can be used to help choose good starting geometries and Hessians to ensure rapid convergence to equilibrium geometries. However, for transition states there is only the vague notion that they must be somewhere between reactants and products. Various strategies have been devised to obtain a sufficiently good initial guess for the transition structure so that quasi-Newton methods will converge.

One approach that has been used for optimizing transition states is to turn the saddle point search into a minimization by optimizing[44,45] the gradient norm, $|\mathbf{g}| = \sqrt{\mathbf{g}^t\mathbf{g}}$. Since the gradient is zero at saddle points as well as minima, minimization of the gradient norm will lead directly to the saddle point, if the starting structure is within the quadratic region to the saddle point. This works well for some transition states, but the radius of convergence is even smaller than conventional quasi-Newton methods modified for transition structure searching. Figure 4 shows the gradient norm for the surface in Fig. 1, and it is clear that the topology of the gradient norm surface is much more complicated and the optimization is more problematic than on the original potential energy surface. The difficult task is still to get within the quadratic region of the transition structure so that the optimization will converge.

For very small systems, it may be feasible to use a grid search to locate the quadratic region of the saddle point. Provided that only 3–5 points are calculated along each coordinate, the total number of points that must be calculated is not unreasonable if the search is limited to three or four dimensions. However, many transition states involve coupled motion of more than three or four coordinates and thus cannot be explored economically by calculating a grid of energies.

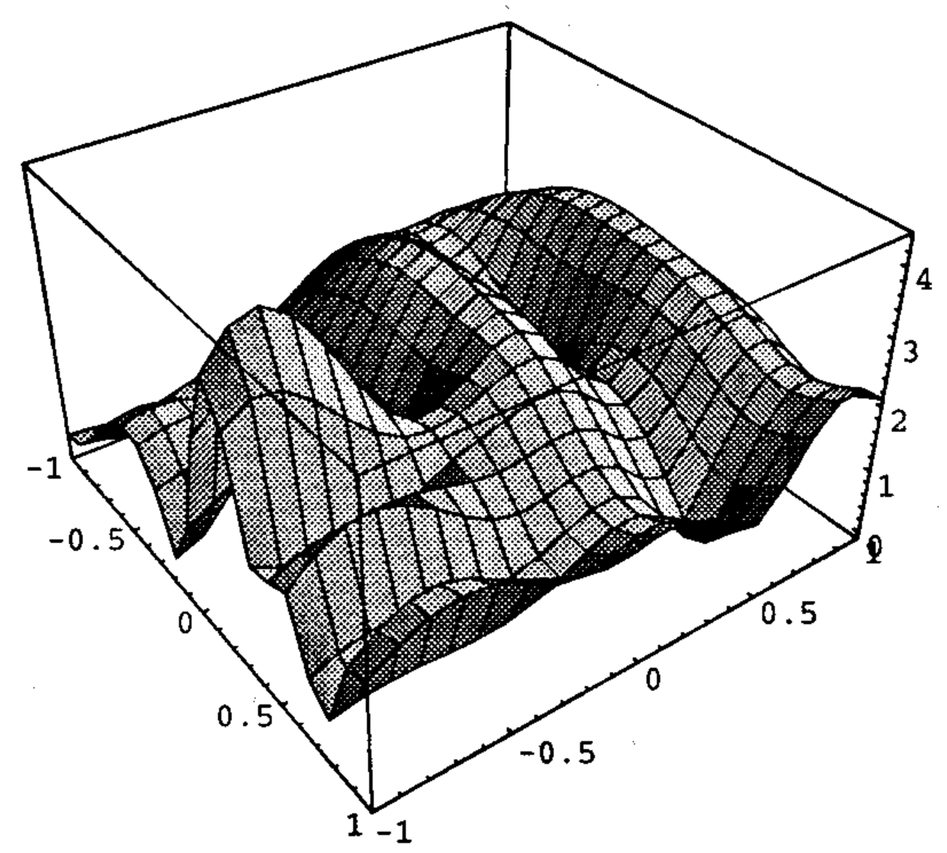


Fig. 4. Gradient norm for the model potential energy surface shown in Fig. 1.

4.1. *Linear and Quadratic Synchronous Transit*

If the geometries of the reactants and products are known, then a rough approximation to the reaction path is a linear interpolation between these two structures.[46] The maximum along this path is then a crude approximation to the transition structure, as shown in Fig. 5. The interpolation can be done in Cartesian coordinates, internal coordinates or distance matrix coordinates (i.e., all $n(n-1)/2$ interatomic distances in an n-atom molecule). The latter has the advantage that it preserves the structural relations between parts of the molecules that do not change during the course of the reaction. The maximum along the linear synchronous transit path[46] (LST) gives an upper bound to the energy barrier for the reaction. Although this structure frequently has the wrong number of negative eigenvalues of

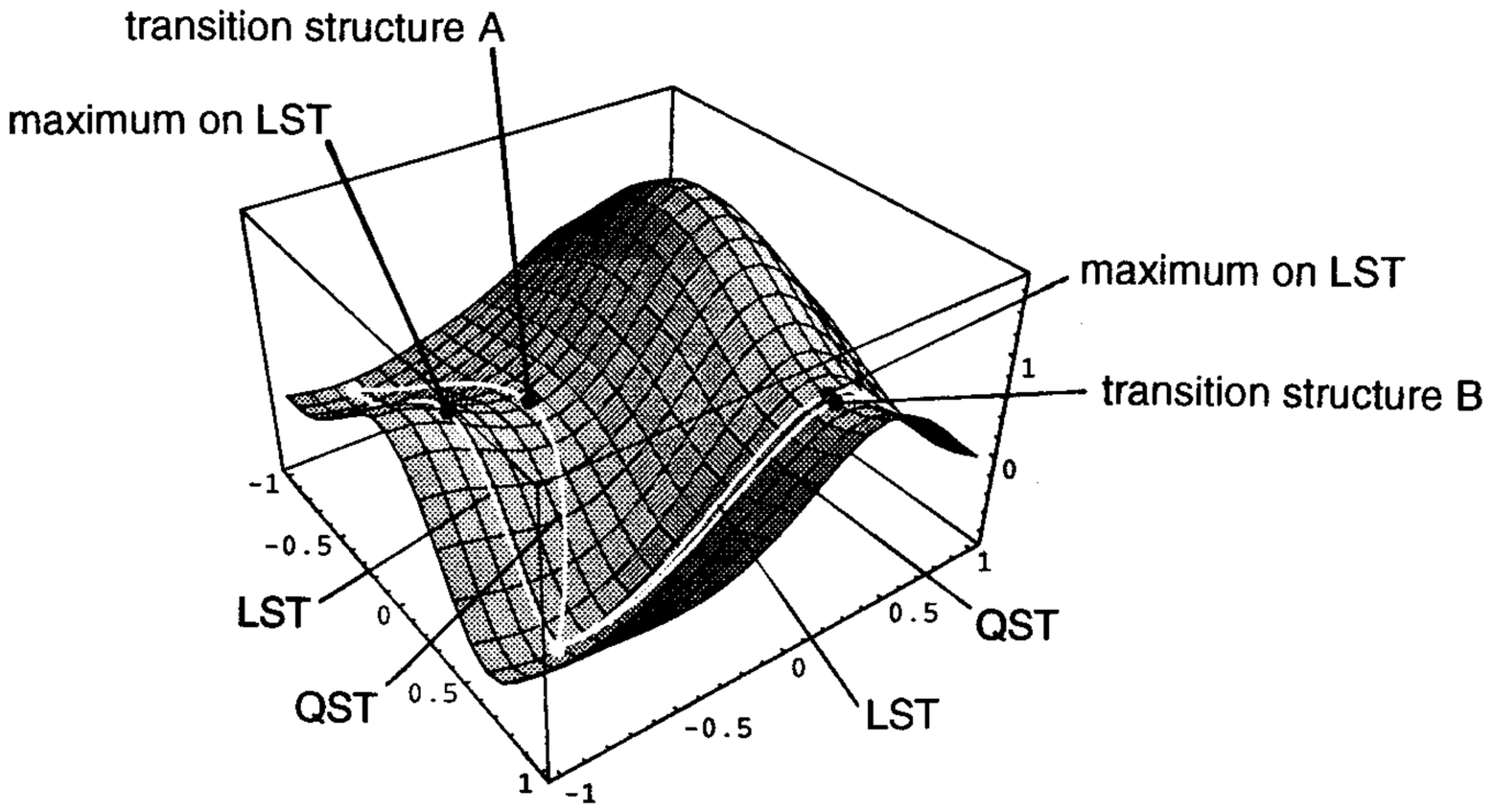


Fig. 5. A model potential energy surface[25] illustrating the linear synchronous transit (LST) and quadratic synchronous transit (QST) methods.

the Hessian (e.g., the LST between reactants and product A in Fig. 5), it is often close enough to the transition structure that a quasi-Newton optimization will succeed.

An improvement on the LST approach can be achieved by using a quadratic synchronous transit path[46] (QST). A lower bound to the reaction barrier can be obtained by minimizing in all directions perpendicular to the LST path. A parabolic interpolation through this point and the reactants and products yields the QST path. The maximum along the QST path is a much better estimate of the transition structure, as indicated in Fig. 5. For higher-dimensional problems, the agreement between the maximum along the first QST path and the true transition structure is not necessarily quite as good, and the process may have to be repeated a number of times. If the reaction path is very curved, or if there are intermediates, then the QST path may also be a poor approximation. This can be remedied in part by using endpoints for the LST and QST paths that are closer to the transition structure. Several gradient-based algorithms have been implemented that alternate between maximizing along an approximate reaction path and minimizing perpendicular to the path.[14,47,48]

4.2. *Coordinate Driving*

For some reactions, a single coordinate can be found that dominates the transformation from reactants to products (e.g., a torsional angle for a conformational transition state or a bond length for a dissociation reaction). An approximate reaction path can be calculated by stepping along the dominant coordinate and minimizing with respect to all $n-1$ remaining coordinates. The maximum along such a path should be quite close to the transition structure and a quasi-Newton based optimization will converge quickly to the saddle point. Since each step along the path requires an $(n-1)$-dimensional optimization, this approach can be rather costly. If a different coordinate becomes dominant after some displacement along the path, there can be abrupt changes in the geometry during the $(n-1)$-dimensional minimization.[49]

4.3. *Walking Up Valleys*

Following the shallowest ascent path, i.e., walking up a valley,[42,50–56] is an alternative to coordinate driving that can reach a saddle point region even when the reaction path is curved. The least steep uphill direction depends on the lowest eigenvector of the Hessian. In many ways, walking up valleys is the converse of following the steepest descent path from a transition state. However, a steepest descent path must end up at a minimum, but a shallowest ascent path need not find the desired transition state, nor even the lowest transition state. As shown in Fig. 6, the search starting from the reactants ends at saddle point B; saddle point A can be reached by walking up the valley from the product side. The walking up valleys approach is usually combined with a quasi-Newton method and is discussed in more detail in Sec. 4.5.

An alternative way to walk up valleys involves approaching the transition state from both sides simultaneously.[57,58] Consider one structure, $\mathbf{R}'$, in the reactant valley and another structure, $\mathbf{P}'$, in the product valley separated by a distance $\delta = |\mathbf{P}' - \mathbf{R}'|$. The lower energy structure is moved to reduce δ by a small amount (5–25%) and the energy for that point is minimized under the constraint that δ remains constant. The process is repeated until δ is sufficiently close to zero (i.e., both structures have converged to the transition structure). Alternatively, once δ is small enough, a quasi-Newton method can be used for the final convergence to the transition structure.[40]

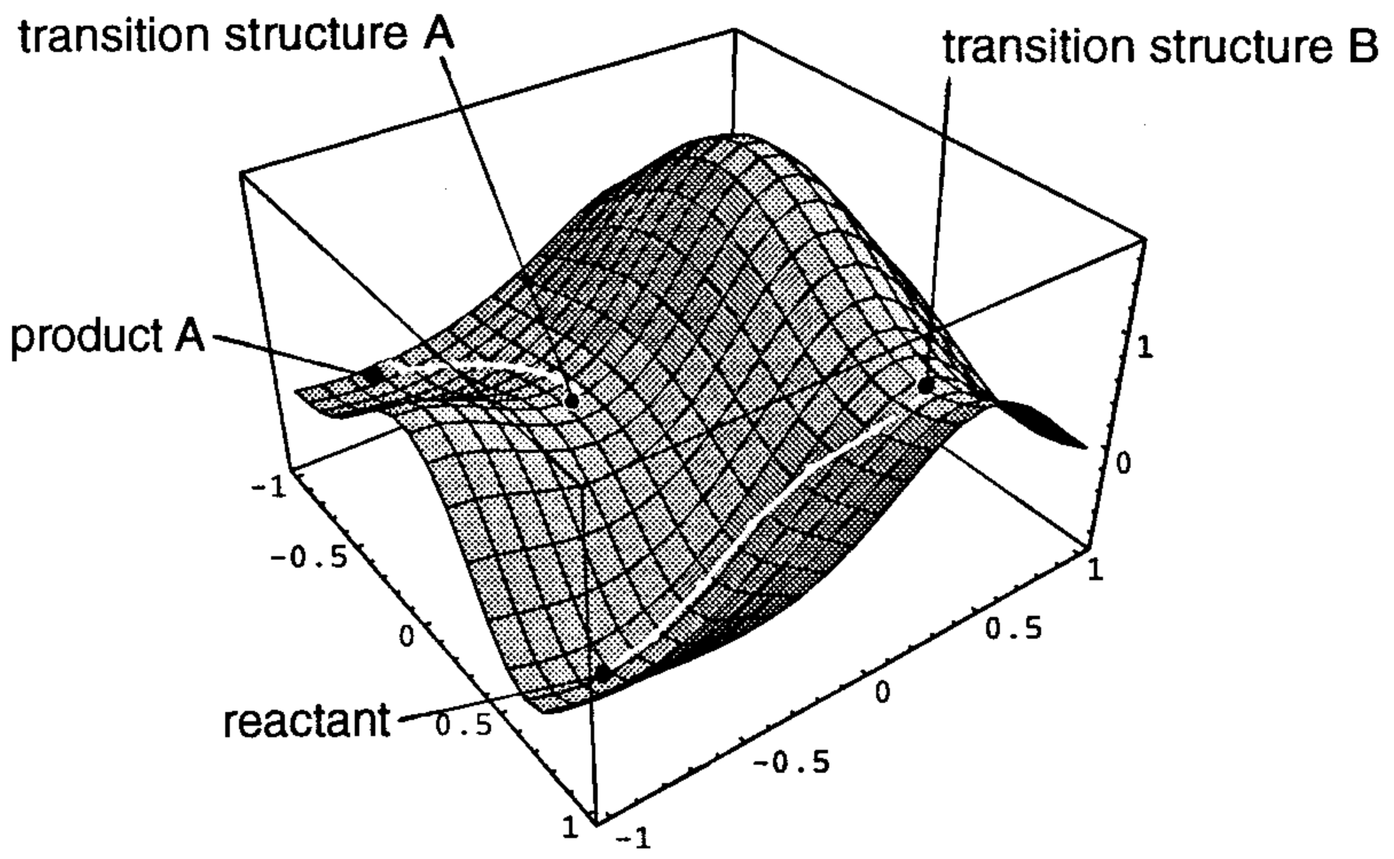


Fig. 6. A model potential energy surface[25] showing a series of points obtained by the walking up valleys or eigenvector following method.

4.4. *Reaction Path Approach*

A more flexible representation of an approximate reaction path might involve several points interpolated between reactants and products. The positions of these points can be refined to improve the initial estimate of the reaction path. The maximum along the best path is a very good estimate of the transition structure. Elber and Karplus[59] have developed this approach and used it to find transition states in macromolecular systems by molecular mechanics methods; a similar method has been outlined by Stacho and Bán.[60] A number of points are spaced equally along an initial reaction path. The potential energies of these points are minimized with the constraint that the spacing between the points remains equal (however, the overall length of the path, L, can change):

minimize $S(\mathbf{x}_0, \mathbf{x}_m)$, where

$$S = \frac{1}{L}\int_{\mathbf{x}_0}^{\mathbf{x}_m} E(\mathbf{x})ds \approx \frac{1}{L}\sum_{i=1}^{m}\frac{E(\mathbf{x}_i)+E(\mathbf{x}_{i-1})}{2}|\mathbf{x}_i-\mathbf{x}_{i-1}|\,. \quad (32)$$

This path will always go through a saddle point if a barrier exists between reactants and products, but on complicated surfaces, this saddle point is not necessarily the lowest transition structure. Typically 10 or more points are used to represent a path. In the refinement, all of the points on the path are moved, and several hundreds of calculations may be needed for the refinement (a variant that concentrates on the saddle points is somewhat more efficient[59b]). The method will also identify any intermediates between reactants and products, but special care must be taken so that the path does not collapse into the intermediate minima.[59c,d] Although originally formulated for empirical force field calculations, this method has recently been applied to *ab initio* and semi-empirical methods.[61]

4.5. *Quasi-Newton Methods*

Most of the strategies discussed above (LST, QST, coordinate driving, walking up valleys, etc.) locate the general region of the transition state but do not yield a converged saddle point with zero forces. An unconstrained optimization involving all degrees of freedom is needed to complete the search for the transition structure. The quasi-Newton methods discussed in Sec. 3 can be altered to converge on saddle points by modifying the choice of search direction and the control of the step size. A flow chart for a quasi-Newton-based transition state search is shown Fig. 3 and is the same as for minimizations. Some of the strategies discussed above for approaching the saddle point region can also be incorporated into these algorithms. As in Sec. 3, the discussion is divided into updating the Hessian and controlling the step size and direction.

4.5.1. *Hessian Update*

Because a saddle point has one negative eigenvalue, a positive definite Hessian updating method cannot be used. The methods of Powell[35] and Schlegel[38] appear to work well (Eq. (11), Eqs. (15)–(17)). Bofill[62] has found that a combination of the Murtagh–Sargent and symmetric Powell updates is advantageous. The BFGS[36] and DFP[37] methods sometimes work if the initial Hessian has one negative eigenvalue and the step direction is controlled so that an uphill step is taken along the lowest eigenvector. If the Hessian rather than the inverse is updated, it may be convenient to invert the Hessian by diagonalization, so that the size and sign of the eigenvalues can be monitored. If the wrong number of negative eigenvalues is found (e.g., 0 or 2 instead of 1), then the offending values can be reset to the correct sign before computing the

inverse. A better technique is to control the step direction (see below) so that it goes uphill along a suitable direction, and to allow the normal updating procedure to correct the Hessian as the optimization reaches the quadratic region of the saddle point.

4.5.2. *Search Direction and Step Size Control*

Once the structure is within the quadratic region of the saddle point and the Hessian has the right number of negative eigenvalues and appropriate eigenvectors, then quasi-Newton methods will converge to the saddle point. With proper control of the step size and direction, these methods can also converge to the saddle point starting from points outside the quadratic region. The guiding coordinate and eigenvector following methods are discussed below. A variety of other strategies such as Pulay's GDIIS method[29] (Eqs. (23) and (24)) and the image surface method[63,64] are also used successfully by a number of groups.

In the guiding coordinate approach, an approximate reaction path, such as LST or QST, is used to define the direction for maximization. Particularly noteworthy is the method implemented by Bell and Crighton.[14] After a maximum is found along a parabolic path between reactants and products, a minimum is found in the space *conjugate* to the path, rather than *orthogonal* to the path as is done in other implementations. If the surface were quadratic, then the structure would remain at the maximum along the path. Since real surfaces can deviate significantly from quadratic, it is necessary to check whether the structure is still at the maximum at each step. If the gradient along the parabolic path between the reactants, products and the current point is too large, the search for a maximum along the path is repeated. This method works well, provided that the potential energy surface has a maximum along the approximate path.

The eigenvector following approach[42,50–56] is a combination of walking up valleys and quasi-Newton optimization. This is usually developed in terms of the RFO model (Eqs. (29)–(31)), but can also be interpreted in terms of trust radius restriction of the step size (Eqs. (26)–(27)). As in the case of minimization, it is best to work in terms of the eigenvalues and eigenvectors of the Hessian (b_i and $\mathbf{v}_i$ respectively, where $b_1 < b_2 <$ etc.). A step is taken uphill along the lowest eigenvector and downhill along all other eigenvectors. Based on several years of experience with this approach,

Simons *et al.*[54,55] recommend:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha \sum_{j=1}^{n} \frac{\mathbf{v}_j(\mathbf{v}_j^t \mathbf{g}_i)}{(b_i - \lambda)}\,,$$

$$\lambda = (2b_1 + b_2)/4\,, \text{ and } \alpha = 1 \text{ for } 0 < b_1 < b_2/2\,, \tag{33}$$

$$\lambda = (b_1 + b_2)/4\,, \text{ and } \alpha = 1 \text{ for } b_1 < 0\,,$$

$$\lambda = (3b_1 + b_2)/4\,, \text{ and } \alpha = (b_2 - b_1)/b_1 \text{ if } b_1 > b_2/2\,.$$

Alternatively, the problem can be partitioned into a maximization and a minimization, and different step lengths used in the two spaces.[42,56] Let λ_0 control the uphill step and λ the downhill step. The step is given by

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha \left(\frac{\mathbf{v}_1(\mathbf{v}_1^t \mathbf{g}_i)}{(b_1 - \lambda_0)} + \sum_{j=2}^{n} \frac{\mathbf{v}_j(\mathbf{v}_j^t \mathbf{g}_i)}{(b_i - \lambda)} \right)\,. \tag{34}$$

So that the step is in the uphill direction along $\mathbf{v}_1$, λ_0 must be greater than b_1; λ_0 is chosen by solving the RFO problem (Eq. (30)) for this direction:

$$\begin{bmatrix} b_1 & \mathbf{v}_1^t \mathbf{g}_i \\ \mathbf{v}_1^t \mathbf{g}_i & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^t \Delta\mathbf{x} \\ 1 \end{bmatrix} = \lambda_0 \begin{bmatrix} \mathbf{v}_1^t \Delta\mathbf{x} \\ 1 \end{bmatrix}\,,$$

$$\lambda_0 = \frac{1}{2} b_1 + \frac{1}{2} \sqrt{b_1^2 + 4(\mathbf{v}_1^t \mathbf{g}_i)^2}\,. \tag{35}$$

The value of λ for the downhill step is found by solving the RFO problem for the remaining $(n-1)$-dimensional space. If the step is longer than the trust radius, it is scaled back by adjusting α in Eq. (34). Culot *et al.*[65] have devised a third alternative; the formula is similar to Eq. (34) but with $\lambda_0 = \lambda'$ and $\lambda = -\lambda'$, where λ' is chosen so that $(b_1 - \lambda') > 0$ and $(b_i + \lambda') > 0$ for $i > 1$, and so that the resulting step lies within the trust radius. In all transition structure optimizations of this type, shorter steps are preferred[54,56] to prevent "stitching" (zig-zag steps across the valley). However, some caution is needed with the trust radius updating so that it does not become unreasonably small.

If a mode other than the lowest eigenvector needs to be followed, the simplest approach is to rescale the variables so that the desired mode becomes the lowest eigenvector.[53–56] To follow mode $\mathbf{v}_k$, the new displacement

$\Delta\mathbf{x}^*$, gradient $\mathbf{g}^*$ and Hessian eigenvalues b_i become:

$$\begin{aligned} \mathbf{v}_k^t\Delta\mathbf{x}^* = \Delta\mathbf{x}/\beta \quad &\text{and} \quad \mathbf{v}_j^t\Delta\mathbf{x}^* = \Delta\mathbf{x},\ j \neq k\,, \\ \mathbf{v}_k^t\mathbf{g}^* = \beta\mathbf{g} \quad &\text{and} \quad \mathbf{v}_j^t\mathbf{g}^* = \mathbf{g},\ j \neq k\,, \\ b_k^* = \beta^2 b_k \quad &\text{and} \quad b_j^* = b_j,\ j \neq k\,. \end{aligned} \tag{36}$$

Either of the above approaches to eigenvector following can then be used without further modification.

5. Algorithms for Finding Surface Intersections and Points of Closest Approach

The lowest point on the seam of intersection between surfaces $E^I(\mathbf{x})$ and $E^J(\mathbf{x})$ can be found by a constrained minimization of the energy[66]:

$$\text{minimize } E^I(\mathbf{x}) \text{ subject to } \Delta E^{IJ} = E^I(\mathbf{x}) - E^J(\mathbf{x}) = 0\,. \tag{37}$$

If the Lagrangian is written as $L = E^I + \lambda\Delta E^{IJ}$, and the potential energy surfaces are expanded as quadratic functions, then the conditions $\partial L/\partial\mathbf{x} = 0$ and $\partial L/\partial\lambda = 0$ lead to the following matrix equation:

$$\begin{bmatrix} \mathbf{H}^I + \lambda\mathbf{H}^{IJ} & \mathbf{g}^{IJ} \\ \mathbf{g}^{IJt} & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \delta\lambda \end{bmatrix} = - \begin{bmatrix} \mathbf{g}^I + \lambda\mathbf{g}^{IJ} \\ \Delta E^{IJ} \end{bmatrix}, \tag{38}$$

where $\mathbf{g}^I = \partial E^I/\partial\mathbf{x}$, $\mathbf{g}^{IJ} = \partial\Delta E^{IJ}/\partial\mathbf{x}$, $\mathbf{H}^I = \partial^2 E^I/\partial\mathbf{x}^2$ and $\mathbf{H}^{IJ} = \partial^2\Delta E^{IJ}/\partial\mathbf{x}^2$. These derivatives can be computed efficiently using density matrix techniques.[20] The necessary code to solve these equations by standard quasi-Newton methods has been implemented by several groups.[66–69] Note that the energies and gradients of the two surfaces must be computed by comparable means, e.g., state-averaged MCSCF.

The simplest way to find the point of closest approach between two surfaces is to carry out an unconstrained minimization on the energy difference between the surfaces, ΔE^{IJ}. A conical intersection or a weakly avoided touching of surfaces can be approximated by two quadratic surfaces, E_0^I and E_0^J, interacting via a matrix element h:

$$\begin{aligned} E &= \left[E_0^I + E_0^J \pm \sqrt{(E_0^I - E_0^J)^2 + 4h^2}\right]/2\,, \\ \Delta E^{IJ} &= \sqrt{(E_0^I - E_0^J)^2 + 4h^2}\,, \\ \frac{\partial\Delta E^{IJ}}{\partial\mathbf{x}} &= \left((E_0^I - E_0^J)\frac{\partial(E_0^I - E_0^J)}{\partial\mathbf{x}} + 4h\frac{\partial h}{\partial\mathbf{x}}\right) \Big/ \sqrt{(E_0^I - E_0^J)^2 + 4h^2}\,. \end{aligned} \tag{39}$$

Some caution is needed for conical intersections, since the gradient and Hessian become undefined at the apex of the cone (i.e., $\partial\Delta E^{IJ}/\partial\mathbf{x} = 0/0$ since $(E_0^I - E_0^J) = 0$ and $h = 0$ at the apex). The gradient and Hessian are better behaved if $(\Delta E^{IJ})^2$ is minimized instead of ΔE^{IJ}. This leads to

$$\begin{aligned} \mathbf{g}^{IJ2} &= \frac{\partial(\Delta E^{IJ})^2}{\partial\mathbf{x}} = 2(E_0^I - E_0^J)\frac{\partial(E_0^I - E_0^J)}{\partial\mathbf{x}} + 8h\frac{\partial h}{\partial\mathbf{x}}\,, \\ \mathbf{H}^{IJ2} &= \frac{\partial^2(\Delta E^{IJ})^2}{\partial\mathbf{x}\partial\mathbf{y}} \\ &= 2(E_0^I - E_0^J)\frac{\partial^2(E_0^I - E_0^J)}{\partial\mathbf{x}\partial\mathbf{y}} + 8h\frac{\partial^2 h}{\partial\mathbf{x}\partial\mathbf{y}} \\ &\quad + 2\frac{\partial(E_0^I - E_0^J)}{\partial\mathbf{x}}\frac{\partial(E_0^I - E_0^J)}{\partial\mathbf{y}} + 8\frac{\partial h}{\partial\mathbf{x}}\frac{\partial h}{\partial\mathbf{y}}\,, \\ \Delta\mathbf{x} &= -(\mathbf{H}^{IJ2})^{-1}\mathbf{g}^{IJ2}\,. \end{aligned} \tag{40}$$

Standard quasi-Newton methods can be used to solve these equations but, as noted above, the two surfaces must be computed by comparable means. This approach has been implemented by a number of groups.[70–72] To avoid using Lagrangian multipliers, the quantity $(\Delta E^{IJ})^2$ should be minimized in the space spanned by $\partial\Delta E^{IJ}/\partial\mathbf{x}$ and $\partial h/\partial\mathbf{x}$, and E^I minimized in the remainder of the space.[72]

6. Algorithms for Following Reaction Paths

One method of following the reaction path is to use the gradient repeatedly to take small steps downhill:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - s\mathbf{g}_i/|\mathbf{g}_i|\,, \tag{41}$$

where s is the step length. This is equivalent to using the Euler method to solve the differential equation for the reaction path, Eq. (1). Unfortunately, Eq. (1) is a stiff differential equation and very small steps are needed to integrate it by the Euler method. As shown in Fig. 7, a step along the gradient deviates from the reaction path if the path is curved. In the reaction path following method of Ishida, Morokuma, Komornicki[73] (Fig. 7(a)), a displacement back to the path is obtained by minimizing the energy along $(\mathbf{g}_i/|\mathbf{g}_i| - \mathbf{g}_i^*/|\mathbf{g}_i^*|)/2$, where $\mathbf{g}_i^*$ is the gradient at the end of the Euler step.

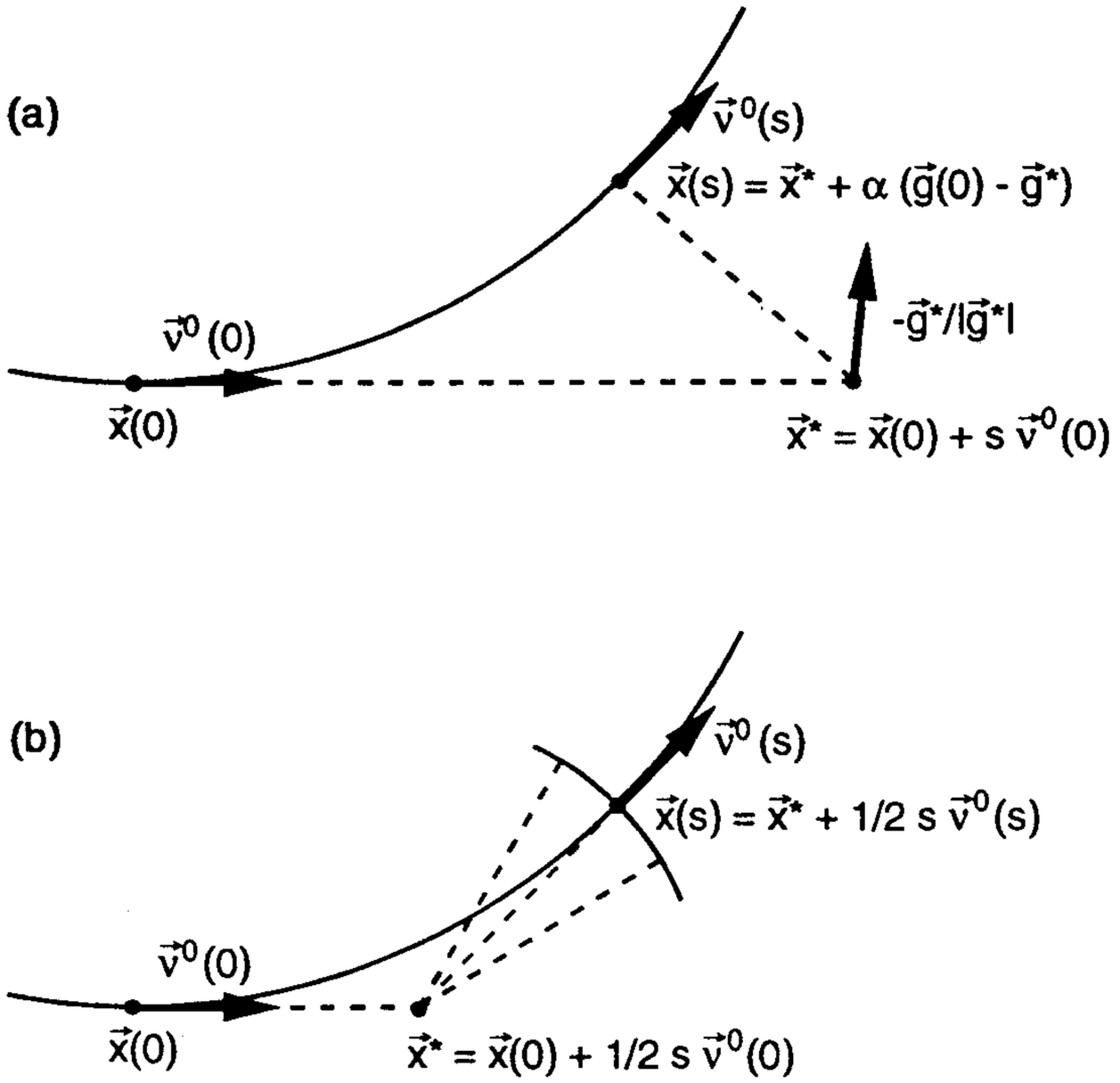


Fig. 7. Reaction path following methods of (a) Ishida, Morokuma and Komornicki (Ref. 73) and (b) Gonzalez and Schlegel (Ref. 77(a)) (solid curves are the approximate paths, arrows indicate the tangents and point downhill).

This is also known as the stabilized Euler method and performs significantly better than the simple Euler method (note, however, in the limit of very small step size, the stabilization must be omitted to get the proper behavior).

Higher-order methods for integrating differential equations, such as fourth-order Runge–Kutta and predictor–corrector methods, are better than Euler and stabilized Euler,[74] but must also be used with small step sizes because the defining equation for the reaction path is inherently a stiff differential equation.

Page and McIver[75] developed a second-order method for following the reaction path by using a quadratic function for the energy to expand the gradient in Eq. (1).

$$\frac{\partial \mathbf{x}(s)}{\partial s} = \frac{-\mathbf{g}_i - \mathbf{H}(\mathbf{x}(s) - \mathbf{x}_i)}{|\mathbf{g}_i + \mathbf{H}(\mathbf{x}(s) - \mathbf{x}_i)|}\,. \tag{42}$$

In the LQA method (locally quadratic approximation), this expression is integrated numerically for a fixed arc length to get the next point along the reaction path. Sun and Ruedenberg[76] have improved the accuracy of the LQA approximation by using a quadratic expansion of the potential energy surface about the center of the step, rather than the beginning of the step. The LQA approach can be corrected by adding the cubic terms in the energy surface (CLQA method)

$$\mathbf{x}_{i+1,\mathrm{CLQA}} = \mathbf{x}_{i+1,\mathrm{LQA}} + \frac{1}{6}\mathbf{v}^2 \Delta s^3\,; \quad \mathbf{v}^2 = -(\mathbf{F}' - (\mathbf{v}^{0t}\mathbf{F}'\mathbf{v}^0)\mathbf{I})\mathbf{v}^0/|\mathbf{g}|\,, \tag{43}$$

where $\mathbf{v}^2 = \partial \mathbf{v}^1/\partial s$ is the rate of change in the curvature, $F'_{ij} = \Sigma_k F_{ijk} v^0_k$ and F_{ijk} are the third derivatives. A family of higher order methods can be developed along the same lines by using higher order polynomial functions for the energy and the corrections.[75]

Gonzalez and Schlegel[77] constructed a different second order method by representing a segment of the reaction path by an arc of a circle. As shown in Fig. 7(b), an Euler step of length $\frac{1}{2}\ s$ is taken to a pivot point, $\mathbf{x}^*_i$ (no calculation of E or $\mathbf{g}$ at $\mathbf{x}^*_i$). A step of length $\frac{1}{2}\ s$ is taken from the pivot point in the downhill direction and the energy is minimized subject to the step length constraint; this yields the next point on the path, $\mathbf{x}_{i+1}$. The points $\mathbf{x}_i$, $\mathbf{x}^*_i$ and $\mathbf{x}_{i+1}$ form an isosceles triangle that is tangent to the path at $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ by construction. Since two tangents to a circle form an isosceles triangle, the reaction path between $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ can be represented by an arc of a circle. On model surfaces and with larger step sizes, the second order GS method appears to be able to follow the reaction path better than the methods discussed above. This approach is equivalent to the implicit trapezoid method for integrating differential equations (implicit methods are known to be particularly stable for integrating stiff differential equations[78]). A family of higher-order reaction-path-following methods can be constructed in a similar fashion.[77c,d]

Since the gradient is zero at the transition structure, the first step cannot be determined from Eq. (1). The transition vector, i.e., eigenvector of the Hessian associated with the negative eigenvalue, is the tangent to the reaction path at the transition structure. Hence the initial step is along the transition vector. Similarly, the curvature of the reaction path at the transition state cannot be determined from Eq. (2); instead the following must be used:

$$\mathbf{v}^1 = -(\mathbf{H} - 2(\mathbf{v}^{0t}\mathbf{H}\mathbf{v}^0)\mathbf{I})^{-1}(\mathbf{F}' - (\mathbf{v}^{0t}\mathbf{F}'\mathbf{v}^0)\mathbf{I})\mathbf{v}^0\,, \tag{44}$$

where $\mathbf{F}'$ is defined after Eq. (43). Note that the CLQA and GS methods give the correct curvature at the transition state but the Euler, stabilized Euler and LQA methods do not.[75–77]

Variational transition state theory and reaction path Hamiltonian methods require the curvature along the path and the vibrational frequencies perpendicular to the path.[28] These are quite sensitive to the accuracy of the reaction path, and often rather small step sizes must be used to obtain converged results.[74c] Near the transition state, the gradient is small and computing the tangent vector from the gradient, $\mathbf{v}^0 = -\mathbf{g}/|\mathbf{g}|$, is subject to large uncertainties. It may be preferable to compute the tangent from the path, $\mathbf{v}^0 = (\mathbf{x}_{i+1} - \mathbf{x}_{i-1})/|\mathbf{x}_{i+1} - \mathbf{x}_{i-1}|$ (or $\mathbf{v}^0 = (\mathbf{x}^*_i - \mathbf{x}^*_{i-1})/|\mathbf{x}^*_i - \mathbf{x}^*_{i-1}|$ for the GS method). However, this necessitates tightening the convergence criteria used in computing the points along the path when small step sizes are used.

Bifurcations on potential energy surfaces have been studied extensively by Ruedenberg *et al.*[27] One type of bifurcation occurs when a descending valley opens into a ridge and the products lie to either side of the ridge (see Fig. 1). Another type of bifurcation involves an ascending valley that divides and passes through two separate transition states. Each of these bifurcations is characterized by a valley–ridge inflection point, in which an eigenvalue of the Hessian for the degrees of freedom perpendicular to the path changes sign. It should be noted that steepest descent paths (intrinsic reaction coordinates) can bifurcate only at stationary points.[27] When a steepest descent path goes through a valley–ridge inflection point that is not a stationary point, the path does not split in two. However, steepest descent trajectories (or shallowest ascent trajectories, in the case of the second type of bifurcation) started to either side of the path will start to diverge strongly at some distance beyond the valley–ridge inflection point, as shown in Fig. 1. A method for locating valley–ridge inflection points has been implemented by Baker and Gill[79]; frequent evaluations of the Hessian are avoided by using finite differences to monitor the lowest eigenvalue of the Hessian in the space orthogonal to the path.

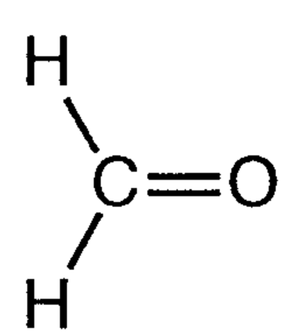
7. Practical Aspects of Optimization

The algorithms outlined above provide some of the mathematical and computational machinery for optimization. The application of these methods to molecular structure problems, however, raises various practical problems

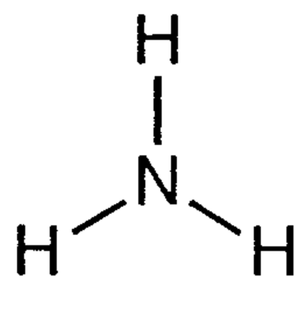
that are not often discussed in the literature. The suggestions summarized in this section are gleaned from a number of years of experience in the optimization of equilibrium geometries and transition structures. Further details can be found in Ref. 18.

7.1. *Symmetry*

The gradient vector belongs to the totally symmetric representation of the point group of the molecule. Thus, if the coordinate system reflects the symmetry of the molecule and round-off errors do not accumulate, then the molecule will not optimize to a lower symmetry, even if a lower symmetry structure is lower in the energy. There are no symmetry restrictions that prevent an optimization from proceeding to a higher symmetry structure if it is lower in energy. If a molecule has been minimized within a particular symmetry, it need not be a minimum with respect to displacements that lower the symmetry. For example, CH_2O optimized under C_{2v} symmetry, **1**, is a minimum with respect to out-of-plane bending, but planar NH_3 optimized under D_{3h} symmetry, **2**, is a maximum with respect to out-of-plane bending. Calculating the full second derivative matrix and diagonalizing it is the best way to test whether an optimized structure is a minimum or saddle point both with respect to displacements within the symmetry and with respect to those that break the symmetry.



1



2

Symmetry reduces the number of coordinates that must be optimized. If a full optimization is desired, then the number of coordinates to optimize is equal to the number of totally symmetric vibrational modes for the molecule. Alternatively, the number of degrees of freedom can be determined by examining the sets of symmetry-equivalent atoms and the symmetry elements. An atom at a center of inversion, the intersection of

two axes or the intersection of an axis or a plane does not contribute to the number of degrees of freedom. A set of symmetry-equivalent atoms contributes one degree of freedom if the atoms are on a proper or improper rotational axes, two degrees of freedom if the atoms are only in a symmetry plane, and three degrees of freedom if they are not on any symmetry element. This total is reduced by the number of rotational and translational degrees of freedom: zero for D_n, D_{nh}, D_{nd} and cubic groups, one for C_{nv} and C_{nh}, two for C_n and S_n, three for C_i and C_s, and six for C_1.

Motion of a molecule on an energy surface can be described by non-rigid symmetry groups and by permutation groups.[80,81] These may be useful in counting up the number of symmetry-equivalent minima, transition states and reaction paths. Stanton and McIver[82] have developed a number of symmetry rules that pertain to transition vectors:

(i) A transition vector with a non-zero imaginary frequency cannot belong to a degenerate representation (if it did, then there would be another, symmetry-equivalent imaginary frequency and the structure would be a higher-order saddle point),

(ii) the transition vector must be symmetric for symmetry operations that leave the reactants and products unchanged, and

(iii) the transition vector must be antisymmetric for symmetry operations that interchange reactants and products.

Examples of these rules are illustrated in Fig. 8. Furthermore, if a transition vector belongs to an antisymmetric representation of the point group, then the transition structure can be optimized by a minimization rather than by a saddle point search (the position of the transition state along the reaction path is determined by symmetry; all that remains is to minimize the rest of the degrees of freedom). Other symmetry and topology rules for transition states have also been discussed.[83]

7.2. *Coordinate Systems*

The choice of coordinates can have a significant effect on the efficiency of a geometry optimization. Most geometry optimizations in *ab initio* molecular orbital theory are carried out in non-redundant internal coordinates (bond lengths, valence angles and dihedral angles), which are very suitable for acyclic molecules. However, considerable care is needed to construct a good set of internal coordinates for cyclic molecules (for a discussion see

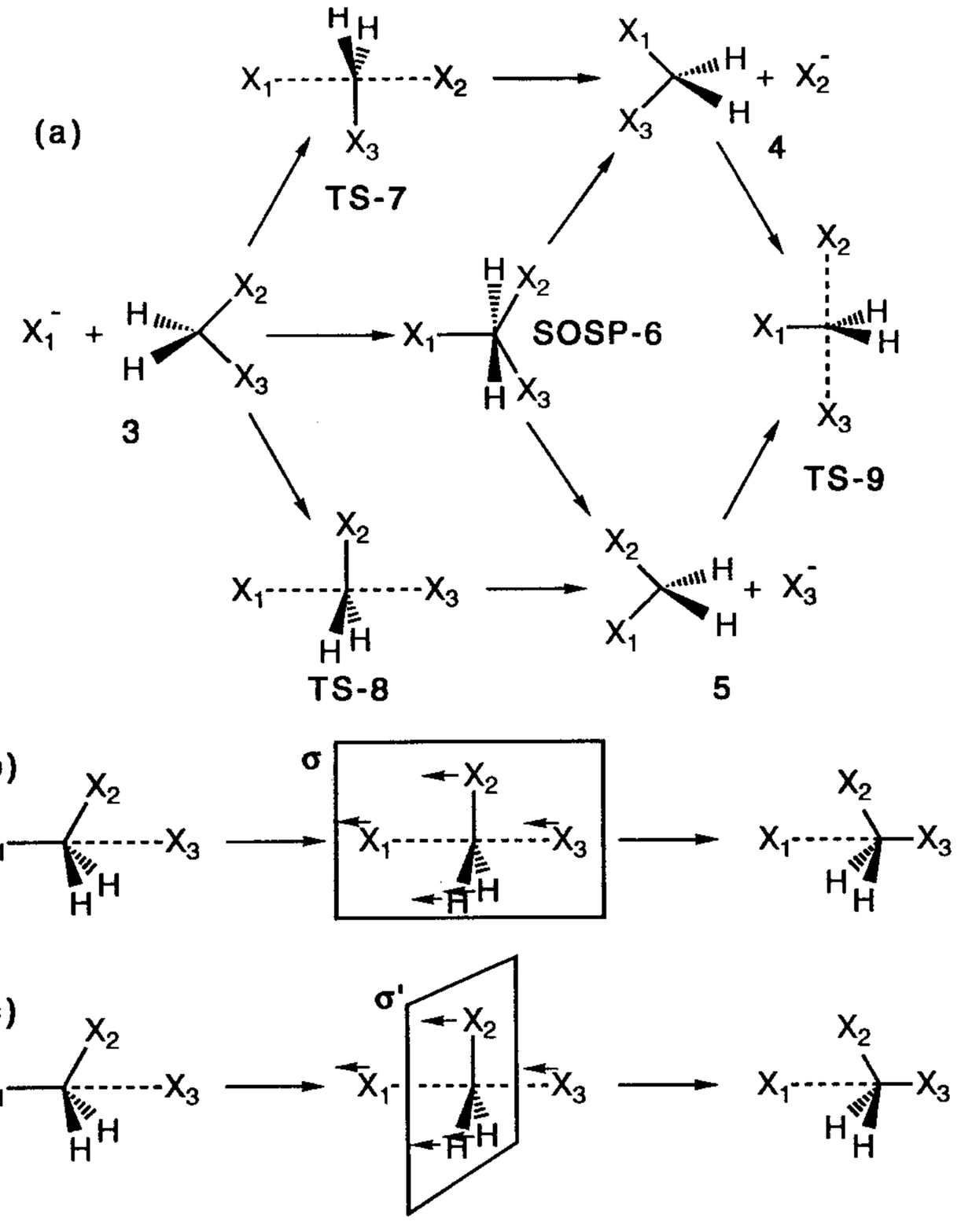


Fig. 8. Symmetry rules for transition states.[82] Consider a degenerate S_N2 reaction with X_1, X_2 and X_3 identical. (a) A transition vector corresponding to a non-zero imaginary frequency cannot belong to a degenerate representation. If **6** were a transition state for **3** → **4**, **3** → **5** and **4** → **5**, then the transition vector would have to be degenerate; however, **6** is a second order saddle point (SOSP), and the actual transition states are **7**, **8** and **9**. (b) A transition vector must be symmetric with respect to symmetry operations that leave reactants and products unchanged. The σ plane shown in the transition state is also a symmetry plane for the reactants and products, and the transition vector is symmetric with respect to this plane. (c) A transition vector must be antisymmetric with respect to symmetry operations that interchange reactants and products. The σ' plane shown in the transition state switches reactants and products, and the transition vector is antisymmetric with respect to this plane. (Figure adapted from S. S. Shaik, H. B. Schlegel and S. Wolfe, *Theoretical Aspects of Physical Organic Chemistry: The S_N2 Mechanism* (John Wiley, New York, 1992)).

Ref. 18). Cartesian coordinates can be better than a poor set of internal coordinates for the optimization of cyclic molecules,[84] provided that a suitable estimate of the starting geometry and the initial Hessian are available. Studies with mixed Cartesian and internal coordinates[85] show that Cartesian coordinates for the rings and internal coordinates for the substituents is more efficient than Cartesian coordinates for all atoms (and that well-chosen non-redundant internal coordinates are better than Cartesian or mixed coordinates). Pulay[86] has found that redundant internal coordinates perform very well for polycyclic molecules, since they display the natural connectivity and curvilinear characteristics of molecular motions.

7.3. *Choice of Starting Geometries and Hessian*

The rate of convergence of an optimization can be accelerated significantly by good choices for the starting geometry and the initial Hessian. Databases containing theoretical[87,88] and experimental[89] structures, semi-empirical calculations and lower level *ab initio* calculations can provide suitable estimates of the geometry, especially if empirically corrected for known systematic errors. An appropriate estimate of the initial Hessian is equally important and can partly compensate for a problematic coordinate system. The diagonal elements of the Hessian should be approximately of the right order of magnitude so that stiff modes are separated from flexible modes, and the off-diagonal elements should express the major couplings between coordinates. For saddle point optimizations, the Hessian should also have one negative eigenvalue, and the corresponding eigenvector should be a suitable approximation to the transition vector. Empirical force fields[90] can provide satisfactory estimates of Hessians for minimizations (even quite simple forms are suitable[91]); however, they must be augmented with additional, nonempirical information for saddle points, since they normally are not parametrized to handle the bond making and breaking that occurs in transition states. Semi-empirical and lower-level *ab initio* calculations yield better estimates of the Hessian for minimizations and saddle point searches, but some caution is needed since potential energy surfaces can change significantly from semi-empirical and low-level *ab initio* calculations to higher-level *ab initio* calculations. For difficult minimizations and especially for problematic transition structure optimizations, it may be necessary to calculate the full Hessian, or at least a few rows and columns of the Hessian, at the same level of theory as used for the optimization.

7.4. *Testing Stationary Points*

After an optimization, it is necessary to test the character of the stationary point by calculating the Hessian at the same level as that used for the optimization. The updated Hessian obtained during the optimization is not sufficiently reliable to confirm the nature of the stationary point within the symmetry and contains no information about displacements that lower the symmetry. If the Hessian cannot be calculated analytically, it can be obtained by numerical differentiation of the gradients. As a side product, the Hessian yields the vibrational frequencies and the zero-point energies; these are needed to compute enthalpy differences, entropies and transition state theory estimates of rate constants. As discussed in Sec. 2, a minimum must have no negative eigenvalues of the Hessian (equivalently, no imaginary frequencies); a first-order saddle point must have one and only one negative eigenvalue (one and only one imaginary frequency). If there are too many negative eigenvalues, a lower energy stationary point of the correct character exists and can usually be found by displacing along the offending eigenvector and re-optimizing (in difficult cases, this may have to be repeated one or more times). If a saddle point optimization ends at a minimum, then the optimization should be restarted after one of the techniques discussed above is used to get closer to the quadratic region of the transition state.

If a transition structure optimization is successful, it is also necessary to examine the transition vector to ensure that it represents a transformation between the desired reactants and products, rather than some unwanted process. If this is not apparent from the nature of the transition vector, it may be necessary to follow the reaction path for some distance down from the transition state in both the forward and reverse directions to determine which reactants and products the transition state connects and to detect whether there are any intermediates along the reaction path. The reaction path (in mass-weighted coordinates) may also be needed for variational transition state theory or reaction path Hamiltonian calculations of the reaction rate.[28]

8. Summary

The preceding pages have served to outline various aspects of optimization applied to the exploration of molecular potential energy surfaces. A number of methods for finding minima, transition structures, surface crossings,

funnels and reaction paths have been discussed. This is intended to be a survey of some of the more efficient or widely used algorithms for optimization tasks, rather than a comprehensive review. In order to touch on most of the important topics, the discussion has been kept brief; the reader is encouraged to read the original papers and to study the program listings for further details. Optimization methodology is still evolving; without a doubt there will be improvements to the methods discussed above and hopefully there will be new algorithms to handle molecular optimization problems that are currently difficult.

Acknowledgments

This work was supported by a grant from the National Science Foundation (CHE 90-20398). Collaboration with Gaussian Inc. on optimization problems is also gratefully acknowledged.

References

1. P. Pulay, Chapter 19 in Part II of this book; R. Shepard, Chapter 7 in this book.
2. P. Pulay, *Adv. Chem. Phys.* **69**, 241 (1987).
3. *Geometrical Derivatives of Energy Surfaces and Molecular Properties*, eds. P. Jørgensen and J. Simons, (D. Reidel, Dordrecht, 1986).
4. T. Helgaker and P. Jørgensen, *Adv. Quantum Chem.* **19**, 183 (1988).
5. For some useful numerical analysis texts on non-linear optimization methods, see Refs. 6–11.
6. R. Fletcher, *Practical Methods of Optimization* (John Wiley, Chichester, 1981).
7. P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization* (Academic, New York, 1981).
8. *Non-linear Optimization, 1981*, ed. M. J. D. Powell (Academic, New York, 1982).
9. J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Non-Linear Equations* (Prentice-Hall, Englewood Cliffs, 1983).
10. L. E. Scales, *Introduction to Non-linear Optimization* (Macmillan, Basingstoke, 1985).
11. W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes* (Cambridge Univ. Press, Cambridge, 1989).
12. For reviews and leading references on optimization of equilibrium geometries and searching for transition states, see Refs. 13–18.
13. K. Müller, *Angew. Chem. Int. Edn. Engl.* **19**, 1 (1980).

14. S. Bell and J. S. Crighton, *J. Chem. Phys.* **80**, 2464 (1984).
15. J. D. Head, B. Weiner and M. Zerner, *Int. J. Quantum Chem.* **33**, 177 (1985).
16. D. A. Liotard, *Int. J. Quantum Chem.* **44**, 723 (1992).
17. H. B. Schlegel, *Adv. Chem. Phys.* **67**, 249 (1987).
18. H. B. Schlegel, in *New Theoretical Concepts for Understanding Organic Reactions*, ed. J. Bertrán, (Kluwer Academic, Dordrecht, 1989), NATO-ASI series C 267, pp. 33–53.
19. D. R. Yarkony, Chapter 11 in this book.
20. B. H. Lengsfield III and D. R. Yarkony, *Adv. Chem. Phys.* **82**(2), 1 (1992).
21. J. Michl and V. Bonačić-Koutecký, *Electronic Aspects of Organic Photochemistry* (Wiley-Interscience, New York, 1990).
22. M. D. Newton, *Chem. Rev.* **81**, 767 (1991); V. Sidis, *Adv. Chem. Phys.* **82**(2), 73 (1992).
23. D. A. McQuarrie, *Statistical Thermodynamics* (Harper & Row, New York, 1973).
24. J. I. Steinfeld, J. S. Fracisco and W. L. Hase, *Chemical Kinetics and Dynamics*, (Prentice-Hall, Englewood Cliffs, 1989).
25. For the equations for the surfaces, see the note on page 500. Surfaces were drawn with *Mathematica* (S. Wolfram, *Mathematica* (Addison-Wesley, Redwood City, 1988) and associated computer programs).
26. K. Fukui, *Acc. Chem. Res.* **14**, 363 (1981).
27. P. Valtazanos and K. Ruedenberg, *Theoret. Chim. Acta* **69**, 281 (1986).
28. D. G. Truhlar and B. C. Garrett, *Acc. Chem. Res.* **13**, 440 (1980); W. H. Miller, N. C. Handy and J. E. Adams, *J. Chem. Phys.* **72**, 99 (1980); D. G. Truhlar and M. S. Gordon, *Science* **249**, 491 (1990).
29. D. K. Hoffman, R. S. Nord and K. Ruedenberg, *Theoret. Chim. Acta* **69**, 265 (1986); J.-Q. Sun and K. Ruedenberg, *J. Chem. Phys.* **98**, 9707 (1993).
30. P. J. M. Van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications* (D. Reidel, Dordrecht, 1988).
31. C. G. Broyden, *Math. Comp.* **21**, 368 (1967).
32. B. A. Murtagh and R. W. H. Sargent, *Comput. J.* **13**, 185 (1972).
33. R. Fletcher and C. M. Reeves, *Comput. J.* **7**, 149 (1964).
34. E. Polak, *Computational Methods in Optimization: A Unified Approach* (Academic, New York, 1971).
35. M. J. D. Powell, in *Nonlinear Programming*, eds. J. B. Rosen, O. L. Mangasarian and K. Ritter (Academic, New York, 1970).
36. R. Fletcher and M. J. D. Powell, *Comput. J.* **6**, 163 (63); W. Davidon, Argonne National Lab. Report ANL-5900.
37. C. G. Broyden, *J. Inst. Math. Appl.* **6**, 76 (1970); R. Fletcher, *Comput. J.* **13**, 317 (1970); D. Goldfarb, *Math. Comput.* **24**, 23 (1970); D. F. Shanno, *Math. Comput.* **24**, 647 (1970).
38. H. B. Schlegel, *J. Comput. Chem.* **3**, 214 (1982).
39. P. Császár and P. Pulay, *J. Mol. Struct.* **114**, 31 (1984).
40. P. L. Cummins and J. E. Gready, *J. Comp. Chem.* **10**, 939 (1989).

41. J. M. McKelvey and J. F. Hamilton, Jr., *J. Chem. Phys.* **80**, 579 (1984).
42. A. Banerjee, N. Adams, J. Simons and R. Shepard, *J. Phys. Chem.* **89**, 52 (1985).
43. J. L. M. Dillen, *J. Comput. Chem.* **8**, 1099 (1987), D.-H. Lu, M. Zhao and D. G. Truhlar, *J. Comput. Chem.* **12**, 376 (1991); T. Schlick, C. Peskin, S. Broyde and M. Overton, *J. Comput. Chem.* **8**, 1199 (1987), and references therein.
44. J. W. McIver, Jr. and A. Komornicki, *J. Am. Chem. Soc.* **94**, 2625 (1972); A. Komornicki, K. Ishida, K. Morokuma, R. Ditchfield and M. Conrad, *Chem. Phys. Lett.* **45**, 595 (1977).
45. D. Poppinger, *Chem. Phys. Lett.* **35**, 550 (1975).
46. T. A. Halgren and W. N. Lipscomb, *Chem. Phys. Lett.* **49**, 225 (1977).
47. S. Bell, J. S. Crighton and R. Fletcher, *Chem. Phys. Lett.* **82**, 122 (1981).
48. O. Tapia and J. Andrés, *Chem. Phys. Lett.* **109**, 471 (1984).
49. U. Burkert and N. L. Allinger, *J. Comput. Chem.* **3**, 40 (1982).
50. J. Pancíř, *Coll. Czech. Chem. Commun.* **40**, 1112 (1975).
51. C. J. Cerjan and W. H. Miller, *J. Chem. Phys.* **75**, 2800 (1981).
52. M. V. Basilevsky and A. G. Shamov, *Chem. Phys.* **60**, 347 (1981).
53. J. Simons, P. Jørgensen, H. Taylor and J. Ozment, *J. Phys. Chem.* **87**, 2745 (1983).
54. J. Simons and J. Nichols, *Int. J. Quantum Chem., Quantum Chem. Symp.* **24**, 263 (1990).
55. J. Nichols, H. Taylor, P. Schmidt and J. Simons, *J. Chem. Phys.* **92**, 340 (1990).
56. J. Baker, *J. Comput. Chem.* **7**, 385 (1986).
57. K. Müller and L. D. Brown, *Theoret. Chim. Acta* **53**, 75 (1979).
58. M. J. S. Dewar, E. F. Healy and J. J. P. Stewart, *J. Chem. Soc. Faraday Trans. 2* **80**, 227 (1984).
59. R. Elber and M. Karplus, *Chem. Phys. Lett.* **139**, 375 (1987); S. Fischer and M. Karplus, *Chem. Phys. Lett.* **194**, 252 (1992); R. Czerminski and R. Elber, *Int. J. Quantum Chem., Quantum Chem. Symp.* **24**, 167 (1990); R. Czerminski and R. Elber, *J. Chem. Phys.* **92**, 5580 (1990).
60. L. L. Sacho and M. I. Bán, *Theoret. Chim. Acta* **83**, 433 (1992).
61. S. S.-L. Chiu, J. J. W. McDouall and I. H. Hillier, *J. Chem. Soc. Faraday Trans.* **90**, 1575 (1994).
62. J. M. Bofill, *J. Comput. Chem.* **15**, 1 (1994).
63. C. M. Smith, *Theoret. Chim. Acta* **74**, 85 (1988); C. M. Smith, *Int. J. Quantum Chem.* **37**, 773 (1990).
64. T. Helgaker, *Chem. Phys. Lett.* **182**, 503 (1991).
65. R. Culot, G. Dive, V. H. Nguyen and J. M. Ghuysen, *Theoret. Chim. Acta* **82**, 189 (1992).
66. N. Koga and K. Morokuma, *Chem. Phys. Lett.* **119**, 371 (1985).
67. A. Farazdel and M. Dupuis, *J. Comput. Chem.* **12**, 276 (1991).
68. D. R. Yarkony, *J. Am. Chem. Soc.* **114**, 5406 (1992).

69. D. R. Yarkony, *J. Phys. Chem.* **97**, 4407 (1993).
70. D. R. Yarkony, *J. Chem. Phys.* **92**, 2457 (1990); D. R. Yarkony, *J. Phys. Chem.* **94**, 5572 (1990).
71. I. N. Ragazos, M. A. Robb, F. Bernardi and M. Olivucci, *Chem. Phys. Lett.* **197**, 217 (1992).
72. M. J. Bearpark, M. A. Robb and H. B. Schlegel, *Chem. Phys. Lett.* **223**, 269 (1994).
73. K. Ishida, K. Morokuma and A. Komornicki, *J. Chem. Phys.* **66**, 2153 (1977); M. W. Schmidt, M. S. Gordon and M. Dupuis, *J. Am. Chem. Soc.* **107**, 2585 (1985).
74. B. C. Garrett, M. J. Redman, R. Steckler, D. G. Truhlar, K. K. Baldridge, D. Bartol, M. W. Schmidt and M. S. Gordon, *J. Phys. Chem.* **92**, 1476 (1988); K. K. Baldridge, M. S. Gordon, R. Steckler and D. G. Truhlar, *J. Phys. Chem.* **93**, 5107 (1989); V. S. Melissas, D. G. Truhlar and B. C. Garrett, *J. Chem. Phys.* **96**, 5758 (1992).
75. M. Page and J. W. McIver, Jr., *J. Chem. Phys.* **88**, 922 (1988); J. Ischtwan and M. A. Collins, *J. Chem. Phys.* **89**, 2881 (1988); M. Page, C. Doubleday and J. W. McIver, Jr., *J. Chem. Phys.* **93**, 5634 (1990).
76. J.-Q. Sun and K. Ruedenberg, *J. Chem. Phys.* **99**, 5257 (1993).
77. C. Gonzalez and H. B. Schlegel, *J. Chem. Phys.* **90**, 2154 (1989); C. Gonzalez and H. B. Schlegel, *J. Phys. Chem.* **94**, 5523 (1990); C. Gonzalez and H. B. Schlegel, *J. Chem. Phys.* **95**, 5853 (1991); H. B. Schlegel, *J. Chem. Soc. Faraday Trans.* **90**, 1569 (1994).
78. C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations* (Prentice-Hall, Englewood Cliffs, 1971).
79. J. Baker and P. M. W. Gill, *J. Comput. Chem.* **9**, 465 (1988).
80. H. C. Longuet-Higgins, *Mol. Phys.* **6**, 445 (1963); for a review see H. Frei, A. Bauder and H. H. Gunthard, *Top. Curr. Chem.* **81**, 1 (1979).
81. W. G. Klemperer, *J. Chem. Phys.* **56**, 5475 (1972); for additional references see T. J. McLarnan, *Theoret. Chim. Acta* **63**, 195 (1982).
82. R. E. Stanton and J. W. McIver, Jr., *J. Am. Chem. Soc.* **97**, 3632 (1975) and references therein; J. H. Murrell and K. J. Laidler, *Trans. Faraday Soc.* **97**, 1680 (1970).
83. For some leading references, see: P. Mezey, *Potential Energy Hypersurfaces* (Elsevier, Amsterdam, 1976); R. G. A. Bone, *Chem. Phys. Lett.* **193**, 557 (1992); A. Rodger and P. E. Schipper, *Chem. Phys.* **107**, 329 (1986); P. Pechukas, *J. Chem. Phys.* **64**, 1516 (1976).
84. J. Baker and W. J. Hehre, *J. Comput. Chem.* **12**, 606 (1991).
85. H. B. Schlegel, *Int. J. Quantum Chem. Quantum Chem. Symp.* **26**, 243 (1992).
86. P. Pulay and G. Fogarasi, *J. Chem. Phys.* **96**, 2856 (1992); J. Baker, *J. Comput. Chem.* **14**, 1085 (1993).

87. R. A. Whiteside, M. J. Frisch and J. A. Pople, *The Carnegie-Mellon Quantum Chemistry Archive* (3rd Ed, Carnegie-Mellon University, Pittsburgh, 1983) and associated computer database.
88. K. Ohno and K. Morokuma, *Quantum Chemistry Literature Data Base* (Elsevier, Amsterdam, 1982; yearly supplements published in special issues of the journal *Theochem*) and associated computer database.
89. F. H. Allen, O. Kennard and R. Taylor, *Acc. Chem. Res.* **16**, 146 (1983) and associated Cambridge Structural Database.
90. U. Burkart and N. L. Allinger, *Molecular Mechanics*, ACS monograph no. 177, (American Chemical Society, Washington, DC, 1982).
91. H. B. Schlegel, *Theoret. Chim. Acta* **66**, 333 (1984).

Note: The surfaces in Figs. 1, 4–6 are of the form $E = \Sigma_i a_i \exp(-b_i(x - c_i)^2 - d_i(y - e_i)^2)$. For Figs. 1 and 4, $\mathbf{a} = (1.7, 1.7, 0.8, 0.8, -1, -1, -0.25, -0.25, -0.5)$, $\mathbf{b} = (1.5, 1.5, 4, 4, 14, 14, 4, 4, 4)$, $\mathbf{c} = (-0.25, 1.75, -0.1, 1.6, 0.35, 1.15, -0.75, 2.25, 0.75)$, $\mathbf{d} = (4, 4, 4, 4, 4, 4, 4, 4, 4)$, $\mathbf{e} = (0.5, 0.5, -0.95, -0.95, -0.75, -0.75, -0.75, -0.75, 1.2)$. For Figs. 5 and 6, $\mathbf{a} = (1.7, 1.7, 0.8, 0.8, -1, -0.25, -0.25, -0.5)$, $\mathbf{b} = (1.5, 1.5, 4, 4, 8, 2, 2, 4)$, $\mathbf{c} = (-0.25, 1.75, -0.1, 1.6, 0.55, -0.75, 2.25, 0.75)$, $\mathbf{d} = (4, 4, 4, 4, 2, 4, 4, 4)$, $\mathbf{e} = (0.5, 0.5, -0.95, -0.95, -0.75, -0.75, -0.75, 1.2)$. Note that there are other, lower energy saddle points outside of the range $x = \pm 1$, $y = \pm 1$. For Fig. 2, the adiabatic surfaces, $S_\pm = \left[E_+ + E_- \pm \sqrt{(E_+ - E_-)^2 + 4h^2}\right]/2$, are obtained by solving the secular equation for two diabatic surfaces $E_\pm = (x \pm 2)^2 + y^2$ interacting via a matrix element h; in (a) $h^2 = 0$, in (b) $h^2 = 16y^2$, and in (c) $h^2 = 16y^2 + 1$.