Ab Initio Methods in Quantum Chemistry—I
Edited by K. P. Lawley

OPTIMIZATION OF EQUILIBRIUM GEOMETRIES AND TRANSITION STRUCTURES

H. BERNHARD SCHLEGEL*

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

CONTENTS

*Camille and Henry Dreyfus Teacher–Scholar.

I. INTRODUCTION

Over the past 15 years there have been rapid developments in the area of geometry optimization in *ab initio* molecular-orbital calculations. This progress has come about primarily because of energy gradient techniques. Analytical gradient-based optimization methods are almost an order of magnitude faster than optimization algorithms that do not use gradients. As a result, optimizing equilibrium geometries has become almost routine, and finding transition structures has become tractable. This is amply demonstrated in various bibliographies of *ab initio* calculations.[1-3] For example, the *Carnegie–Mellon Archive*[2] contains over 10000 optimized structures. Although there had been a number of earlier calculations of energy derivatives,[4-6] Pulay was the first to demonstrate practical energy gradient computations at the self-consistent field (SCF) level.[7] Now, analytical first derivatives are available for most levels of *ab initio* calculations, and second derivatives are also available for a significant number of methods. Energy derivative methods have been surveyed previously,[8-11] and an up-to-date review is given by P. Pulay elsewhere in this series.[12]

With reasonable computational facilities, full geometry optimizations are feasible for molecules as large as 10 non-hydrogen atoms, and perhaps larger if small basis sets are used. For molecules with four or fewer heavy atoms, current practices in theoretical chemistry make it almost mandatory to carry out full geometry optimization at the SCF level with at least a split valence basis set (plus polarization, if possible). For systems with one and two heavy atoms, geometry optimizations at levels of theory that include electron correlation are rapidly becoming the norm. If possible, all of the coordinates of the molecule should be optimized. In the comparison of energies and geometries of equilibrium structures, transition states or points on a reaction path, subtle but significant changes can occur that may be masked by arbitrarily constraining the optimization to a subset of the variables. In some cases, apparent stationary points found by partial optimization may disappear or change drastically on full optimization (particularly for transition states). Fortunately, full optimization with gradient methods is often not much more expensive than partial optimization. For example, if the positions of the heavy atoms are optimized, the coordinates of the hydrogens can be optimized with little additional effort.

The geometry optimization methods discussed in the present chapter are focused primarily on gradient methods for molecular-orbital calculations. Included under optimization methods are algorithms for finding energy minima, locating transition states and higher-order saddle points, and following reaction paths. Beyond the scope of this chapter, but closely related to geometry optimization, is the topic of energy minimization with respect to the variational parameters of the wavefunction. For example, first and second

derivative methods have greatly improved the rate of convergence of multiconfiguration SCF (MCSCF) calculations. Also not directly addressed by this chapter are optimization methods for empirical force field or molecular mechanics calculations. These usually involve larger molecules with many more variables, but energy functions that can be calculated relatively quickly; the different economics of these calculations requires a somewhat different strategy for efficient geometry optimization.

Section II deals with features of energy surfaces, various definitions and preliminary matters. Section III briefly surveys the calculation of analytical energy derivatives. Section IV discusses some energy minimization algorithms that use the energy only, the first derivatives only, or the first and second derivatives. Section V addresses the more difficult task of locating transition structures. Quite a variety of algorithms can be found, with various ranges of applicability. Section VI examines methods for following reaction paths.

II. FEATURES OF ENERGY SURFACES

The concepts of energy surfaces for molecular motion, equilibrium geometries, transition structures and reaction paths depend on the Born–Oppenheimer approximation to treat the motion of the nuclei separately from the motion of the electrons. Minima on the potential energy surface for the nuclei can then be identified with the classical picture of equilibrium structures of molecules; saddle points can be related to transition states and reaction rates. If the Born–Oppenheimer approximation is not valid, for example in the vicinity of surface crossings, non-adiabatic effects are important and the meaning of classical chemical structures becomes less clear. Non-adiabatic effects are beyond the scope of this chapter and the discussion of energy surfaces and optimization will be restricted to situations where the Born–Oppenheimer approximation is valid.

A. Minima, Maxima and Saddle Points

Minima, maxima and saddle points can be characterized by their first and second derivatives. For a function of several variables, the first derivatives with respect to each of the variables form a vector termed the gradient. The second derivatives form a matrix called the Hessian. In classical mechanics, the first derivative of the potential energy for a particle is minus the force on the particle, and the second derivative (for a quadratic potential) is the force constant. Thus, the negative of the components of the gradient are the forces on the atoms or nuclei in a molecule, and the second derivative matrix, or Hessian, is also called the force-constant matrix.

For a function of one variable, the first derivative of the function is zero at a local minimum or maximum. In more than one dimension, the first derivatives

with respect to all of the variables must be zero, i.e. the gradient vector must have zero length. Equivalently, all of the forces on the atoms in a molecule must be zero; hence such a point is also known as a stationary point. In topology, these points are termed critical points. The nature of the stationary point can be determined from the second derivatives. In one dimension, the second derivative is positive for a minimum, negative for a maximum and zero for a point of inflection. In more than one dimension, the eigenvalues of the second derivative matrix, rather than individual elements of the matrix, determine the characteristics of the stationary point. If all the eigenvalues are positive, the point is a local minimum; if all are negative, it is a local maximum. A (first-order) saddle point or col has one negative eigenvalue and all the rest are positive, i.e. a maximum in one direction and a minimum in all perpendicular directions. Thus a first-order saddle point can correspond to a transition structure (see below). An nth order saddle point has n negative eigenvalues, i.e. is a maximum with respect to n mutually perpendicular directions. A zero eigenvalue indicates a point of inflection for motion along the associated eigenvector.

If at all possible, stationary points found by any optimization method should be characterized by computing the Hessian and examining its eigenvalues. This is especially important for high-symmetry structures (where lower-energy, lower-symmetry structures may exist nearby) and for saddle points (which can be higher-order saddle points rather than first-order).

B. Transition Structures and Intrinsic Reaction Coordinates

A minimum on an energy surface represents an equilibrium structure. If there is more than one minimum on a contiguous energy surface, a family of paths can be constructed that connect one minimum to the other. If the highest-energy point on each path is considered, the transition structure can be defined as the lowest of these maxima, i.e. the top of the barrier for the lowest-energy path from one minimum to another. Thus, the transition structure must be a maximum along the reaction path, and a minimum for all displacements perpendicular to the path. This is just the definition of a first-order saddle point. If a higher-order saddle point were chosen (i.e. maximum in more than one direction), it would be possible to move perpendicular to the path to find another path with a lower maximum. The eigenvector corresponding to the single negative eigenvalue of a first-order saddle point is termed the transition vector. At least initially, the steepest-descent path from the saddle point to either of the two minima follows the transition vector.

Since the minima and saddle point are well defined points on the energy surface, it should be possible to define a unique reaction path. The steepest-descent path from the saddle point to the minima can be defined easily, but depends on the particular choice of coordinate system. Cartesian coordinates would yield a different path than internal coordinates. Furthermore, internal coordinates are not unique, since a number of different sets of bond lengths, angles and torsions can represent the same structure.

An intrinsic reaction path[13] can be defined independently of the coordinate system by appealing to classical mechanics. For a given energy surface, the movement of a classical particle must be the same regardless of whether Cartesian coordinates or any of a number of different sets of internal coordinates are used. The intrinsic reaction coordinate (IRC) can be defined as the path traced by a classical particle sliding with infinitesimal velocity from a saddle point down to each of the minima. Since the classical equations of motion can be defined in any coordinate system, and since they must yield the same trajectory, this definition of the intrinsic reaction coordinate is unique. The classical equations of motion are the simplest in mass-weighted Cartesian coordinates, where the effective mass for each coordinate is unity. In this coordinate system, the intrinsic reaction coordinate is the same as the steepest-descent path. An equivalent definition[14] requires the IRC to be the path of minimum distance (geodesic) between the reactants and products, with the metric determined by the potential energy surface.

Intrinsic reaction coordinates are geometrical or mathematical features of the energy surfaces, like minima, maxima and saddle points. Considerable care should be taken not to attribute too much chemical or physical meaning to the reaction coordinate. Since molecules have more than infinitesimal kinetic energy, a classical trajectory will not follow the intrinsic reaction path and may in fact deviate quite widely from it. The intrinsic reaction coordinate is, however, a convenient measure of the progress of a molecule in a reaction. It also plays a central role in the calculation of reaction rates by variational transition state theory[15] and with reaction path Hamiltonians.[16]

C. Symmetry and Topology

Energy surfaces are smooth functions connecting the minima associated with molecular structures. If there are n minima on one contiguous surface, each minimum must be connected to at least one other minimum (otherwise the surface is not contiguous). Since there must be at least one saddle point between connected minima, there are at least $n-1$ saddle points on a surface with n minima. Other useful relations can also be derived[17] but these either require or yield information about higher-order saddle points or maxima, and are therefore less interesting for chemical reactions.

The energy surface must reflect the symmetry of the molecules it represents. The motion of molecules on energy surfaces has been discussed in terms of non-rigid symmetry groups,[18] and permutation groups have been used to classify reaction paths for isomerization.[19] More directly connected to the present topic are the restrictions that symmetry places on the transition

vectors for reactions.[20,21] Some of the more important symmetry requirements for transition vectors discussed by Stanton and McIver[21] are:

1. The transition vector cannot belong to a degenerate representation. Otherwise there would be at least two symmetry-equivalent eigenvectors of the Hessian with negative eigenvalues and the structure would not be a first-order saddle point or transition state. Closely related to this, it can be shown that three valleys cannot meet at a single transition state (except as a numerical accident).
2. The transition vector must be antisymmetric for the symmetry operations that convert the reactants to products.
3. The transition vector must be symmetric for the symmetry operations that leave the reactants and products unchanged.

If an optimization is required to retain the symmetry of a structure, then symmetry-equivalent parameters must remain equal. Thus only displacements belonging to the totally symmetric irreducible representation are permitted; other displacements are forbidden since they would change the symmetry. If it can be determined (e.g. from the Stanton–McIver symmetry rules) that the transition vector does not belong to the totally symmetric representation, then the position of the transition structure along the reaction path is fixed by symmetry, and only the coordinates perpendicular to the reaction path need to be optimized. Since a transition structure must be a minimum with respect to all displacements other than along the reaction path, optimization of the transition structures is reduced to a simple minimization.

By symmetry, the only components of the gradient that are non-zero are those belonging to the totally symmetric representation. Thus, gradient optimization methods (that update the Hessian symmetrically) will not lead to lower-symmetry structures provided the coordinate system reflects the symmetry of the molecule and numerical round-off errors are negligible. Nevertheless, it is quite possible that distortion to a lower symmetry will reduce the energy. This occurs if the higher-symmetry structure is a local maximum or a saddle point with respect to displacements that do not belong to the totally symmetric representation. Calculation of the full second derivative matrix is the best way to test if an optimized structure is stable or unstable with respect to all possible distortions, including those that lower the symmetry.

Sometimes a molecule moves toward a higher symmetry during an optimization. The convergence to the higher symmetry can be relatively slow, and it may be unclear whether the optimized geometry does have the higher symmetry or whether the molecule is slightly distorted and has a lower symmetry. The problem is best resolved by first optimizing the higher-symmetry structure and then testing whether non-totally symmetric distortions lead to a lower energy, preferably by computing the second derivative matrix (numerically or analytically). If the second derivative matrix of the high-symmetry structure has one or more negative eigenvalues, then the structure is not a local minimum and displacement along an eigenvector that corresponds the negative eigenvalue will lead to a lower-energy, lower-symmetry structure. An optimization on the distorted structure then results in a lower-symmetry, lower-energy stationary point (whose stability with respect to further reduction in the symmetry must also be tested).

D. Choice of Coordinates

Before geometry optimization can be carried out, a coordinate system must be chosen to represent the energy surface and the structure of the molecule. Aside from the algorithm for optimization (see Sections IV and V), the choice of the coordinate system is perhaps the single most important factor in determining the ease or difficulty of an optimization. Internal coordinates are normally preferred for geometry optimization, since the total energy is invariant to overall translation and rotation. Usually, several internal coordinate systems are possible. The coordinates must be non-redundant so that the geometry can be specified uniquely (and without extra constraints), and so that the derivatives can be calculated unambiguously. For example, a tetrahedral center has six valence angles, only five of which are independent; a planar six-membered ring has six bond lengths and six angles, but only nine independent coordinates (examples are discussed in Ref. 22).

Difficulties encountered in optimizations can often be traced to problems caused by the coordinate system, especially in cyclic molecules and in transition structures. A different choice of coordinates may significantly improve the behavior of the optimization. Strong coupling degrades the performance of any optimization method on a non-quadratic surface. Coupling between flexible and stiff modes causes special difficulties. This corresponds to a long narrow valley running diagonally across the energy surface. In pathological cases, the valley is also curved, due to non-quadratic terms in the coupling between the coordinates. It may be possible to find a coordinate system that reduces the curvature and/or the coupling. For transition structure optimization, coupling between the transition vector and the other coordinates is also detrimental. A coordinate system should be selected so that the transition vector is dominated by only a few variables (preferably only one), and so that the coupling of these to the remaining coordinates is small.

III. ANALYTICAL ENERGY DERIVATIVES

A comprehensive review of analytical energy methods has been written by P. Pulay and appears elsewhere in this series.[12] The purpose of this section is

only to outline some of the computational steps and to supply a few leading references.

An electronic wavefunction can be constructed from a linear combination of many-electron configurations, which can be built from one-electron molecular orbitals, which in turn can be formed from linear combinations of basis functions. To compute the energy, the integrals over the basis functions must be calculated and the coefficients must be determined. The integrals are usually computed analytically. Some or all of the linear coefficients are calculated variationally, others may be obtained by non-variational methods; the particular recipe depends on the specific method. For derivatives of the energy, the corresponding derivatives of the integrals over basis functions are always needed. Derivatives of the linear coefficients may or may not be required, depending on whether the coefficients were determined variationally.

For polyatomic calculations, basis sets usually consist of contracted Gaussians, i.e. fixed linear combinations of Gaussian-type primitives. The necessary integrals can be calculated analytically, and a variety of programs have been developed to compute them efficiently.[23,24] The integral derivatives can also be calculated analytically, although with somewhat greater effort than the integrals themselves, since each integral can have up to 12 first derivatives and 78 second derivatives with respect to the function positions. Efficient algorithms have been devised and coded by several groups.[25–37,49,60] Translational invariance can be used to reduce the number to nine first derivatives and 45 second derivatives. Rotational invariance can further reduce the computational work, especially for third derivatives.[38–40,56] Depending on the energy method and whether first, second or higher derivatives are needed, it may be possible to avoid the explicit storage of the integral derivatives and/or the explicit transformation of the integrals.[49,57,58]

For SCF and MCSCF energy calculations, all of the coefficients—molecular-orbital (MO) and configuration-interaction (CI)—are determined variationally, i.e. by minimizing the expectation value of the energy. The derivative of the energy with respect to any change in these coefficients is zero (provided the molecular orbitals remain orthonormal and the CI coefficients remain normalized). Hence the first derivatives of the coefficients are not required for the first derivative of the SCF and MCSCF energy; only the first derivatives of the integrals are needed. Consequently, first derivatives can be calculated fairly easily at the SCF and MCSCF levels. Efficient programs for the SCF gradients have been developed over the last 15 years (see Refs. 8–12). Typically, the time required for all $3N-6$ first derivatives of the internal coordinates of an N-atom molecule is equal to or less than the time needed to compute the total energy (integrals plus SCF). First derivatives of the MCSCF energy have appeared more recently[41–45] (and see Refs. 8–12), and are even more economical relative to the time required for the total energy (primarily because of the length of the MCSCF energy computation). For SCF and MCSCF first derivatives, the integral derivatives do not need to be stored or transformed explicitly.

Analogous to the first derivatives, the second derivatives of the SCF and MCSCF energies do not require the second derivatives of the MO and CI coefficients (provided orthonormality is maintained). However, the first derivatives of the coefficients must be calculated by solving the coupled perturbed Hartree–Fock (CPHF) equations,[6] or the coupled perturbed MCSCF (CPMCSCF) equations. Pople and coworkers[46] developed an efficient method for solving the CPHF equations, which made SCF second derivative calculations practicable for restricted Hartree–Fock (RHF) and unrestricted Hartree–Fock (UHF) wave functions. These methods have been extended to include restricted open-shell SCF calculations.[47,48] More recently, the CPMCSCF equations have been formulated[45,49–52] and MCSCF second derivatives have been implemented by a number of groups.[52–56] The second derivatives of the integrals do not need to be stored; storage of the first derivatives can be avoided for the CPHF equations by constructing the derivatives of the Fock-like matrices as the integral derivatives are computed.[49,57,58] For the CPMCSCF equations, some of the transformed-integral first derivatives appear to be needed.

Third derivatives have also been formulated for the SCF and MCSCF energies by a number of groups[49,59,60] and implemented by at least one group.[60] In addition to the first, second and third derivatives of the integrals, only the first derivatives of the coefficients are needed. This is similar to perturbation theory, where the third-order energy can be computed with the first-order wavefunction.

For energy derivatives at levels of theory other than SCF or MCSCF, the first derivatives of the molecular-orbital coefficients are needed in addition to the derivatives of the integrals over basis functions. The molecular-orbital coefficient derivatives must be determined by solving the appropriate coupled perturbed equations: CPHF for wavefunctions based on a single SCF reference configuration (e.g. configuration-interaction with single and double excitations (CISD), second-order Møller–Plesset theory (MP2), coupled clusters with doubles (CCD), etc.), CPMCSCF for multiple reference based on an MCSCF wavefunction (e.g. multi-reference determinant configuration-interaction (MRDCI)). Because of the special form of the first derivative expression, it is possible to avoid the full CPHF or CPMCSCF equations, and solve only a much smaller set of equations.[61] Energy gradients have been programmed for a number of methods including single-reference CI,[62–64,54] multiple-reference CI,[55] second-order Møller–Plesset theory,[46] third-order Møller–Plesset theory[65] and coupled-cluster methods.[66,67] The latter also requires the solution of the coupled perturbed coupled-cluster equations.

Second derivatives for configuration-interaction, coupled-cluster and many-body perturbation theory require first derivatives of those coefficients determined variationally and the second derivatives of those coefficients not

determined variationally. As with the first derivatives of these methods, the solution of the full second-order coupled perturbed equations for the MO coefficients can be reduced to a much smaller set of equations.[61] Second derivatives have been demonstrated for the CI energy[68] and have been formulated for Møller–Plesset theory and for coupled-cluster methods.[50,69]

IV. LOCATING MINIMA ON ENERGY SURFACES

There are quite a number of methods for unconstrained minimization of non-linear functions of many variables[70–74] and it is difficult to assess the goodness of any specific algorithm quantitatively without considering particular applications. Desirable features of an algorithm include: speed of convergence to the minimum, stability and reliability of the method, and the overall cost of the optimization. The choice of the best method for a particular situation depends on the nature of the function to be minimized, the number of variables, the availability of first (and higher) derivatives, and the cost of evaluating the function and its derivatives. Several representative algorithms are examined in this section, with the special features of *ab initio* calculations and geometry optimization in mind.

For geometry optimizations using quantum-mechanical methods, the cost of the function evaluation, i.e. the energy calculation, is much higher than usually assumed for typical minimization problems in numerical analysis. Thus, the total number of evaluations must be kept to a bare minimum, even at the cost of a more complex algorithm. Almost always, the computer time required by the minimization algorithm will be negligible compared to the time needed for the energy evaluation. For example, the cost of inverting or diagonalizing an $N \times N$ matrix, where N is the number of geometrical parameters, is inconsequential compared to the time to calculate the energy. Exceptions to this include determining MCSCF wavefunctions where the energy must be minimized with respect to 10^3 to 10^5 wavefunction parameters, and empirical force field (molecular mechanics) calculations, where the energy, gradient and second derivatives can be calculated relatively easily (neither of these minimization problems are within the scope of this chapter).

A number of representative methods for locating minima are examined more closely. It is convenient to group the algorithms into three categories, depending on the derivative information used: (a) function value only, (b) function plus first derivatives or gradient, and (c) function plus first and second derivatives.

A. Function-only Methods

Algorithms for minimization that require the evaluation of only the function and not its derivatives can be found in a variety of standard books on

numerical analysis.[70–74] Because derivatives are not available for all levels of theory, these methods have the widest range of applicability, but the penalty is that they also have the slowest convergence. The simplex method and the pattern search method are generally not used for geometry optimization in *ab initio* calculations, since they require too many steps. The sequential univariate search or axial iteration algorithm is perhaps the most frequently used method for geometry optimization. A modified Fletcher–Powell algorithm[75,76] is somewhat faster and has also seen considerable use. These two algorithms are discussed below and are illustrated in Figs. 1 and 2 for a simple quadratic surface. The optimizations are shown for two variables, but it should be kept in mind that real examples will have about 10–20 variables. There are a number of other, perhaps more efficient, algorithms that depend only on the function evaluation. These are treated in the next section since they in fact use derivatives evaluated numerically.

The sequential univariate search or axial iteration method changes one

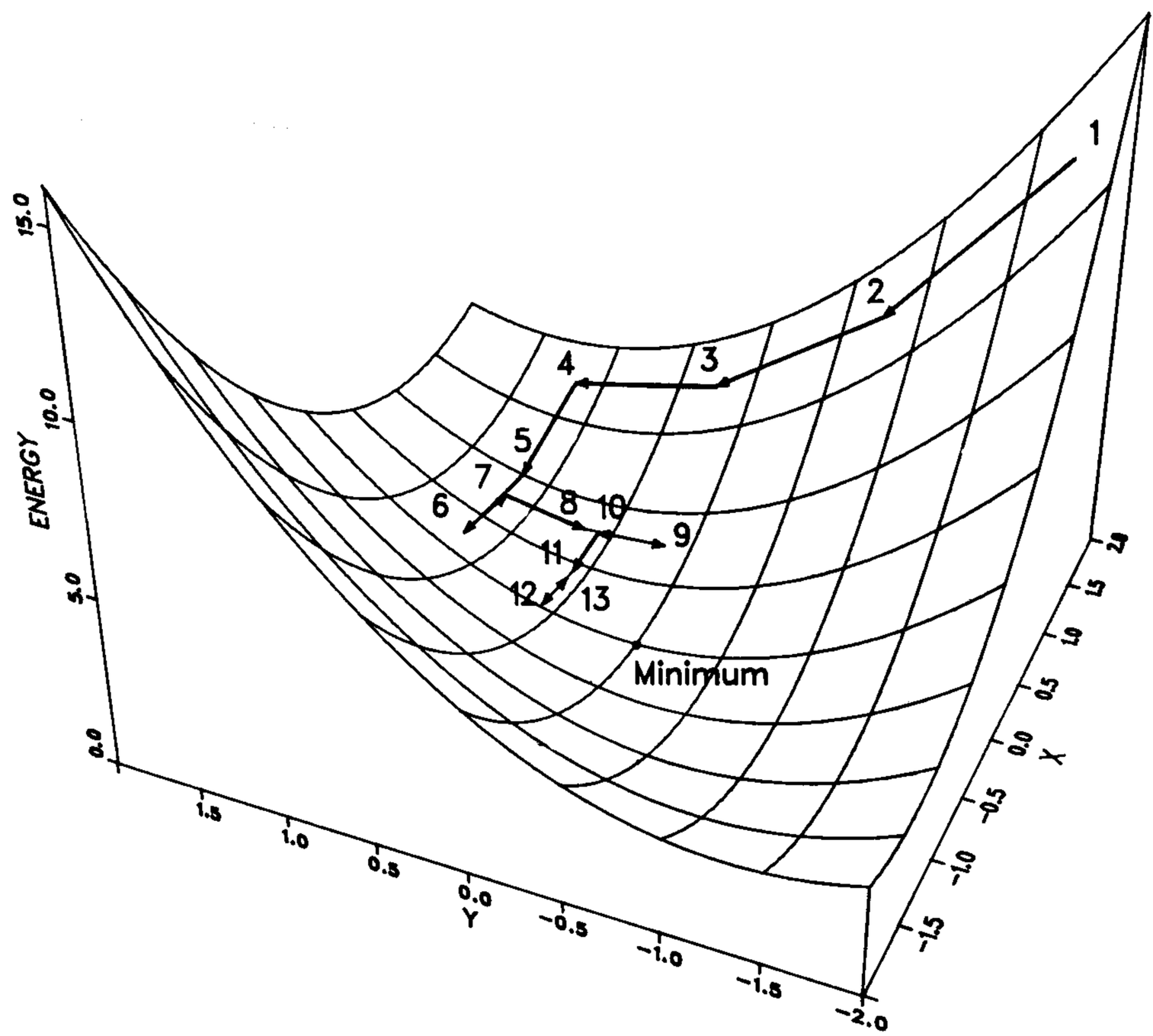

Fig. 1. Steps in finding the minimum on a quadratic surface using the sequential univariate search or axial iteration method.

coordinate at a time (Fig. 1) and cycles over all the coordinates one or more times:

1. Calculate the energy at the initial geometry.
2. Calculate the energy at two displacements along a coordinate.
3. Fit a parabola to the energy at the three geometries (undisplaced and two displaced).
4. Find the minimum of the parabola and calculate the energy at the minimum.
5. Choose the next coordinate and go back to step (2), unless all the coordinates have been processed.
6. If the change in all the coordinates is small enough, then stop; otherwise go back to (2) and cycle through all of the coordinates again.

For the two-dimensional example in Fig. 1, the first cycle gives rise to points (1)–(7). As can be seen, convergence to a minimum is not guaranteed in one

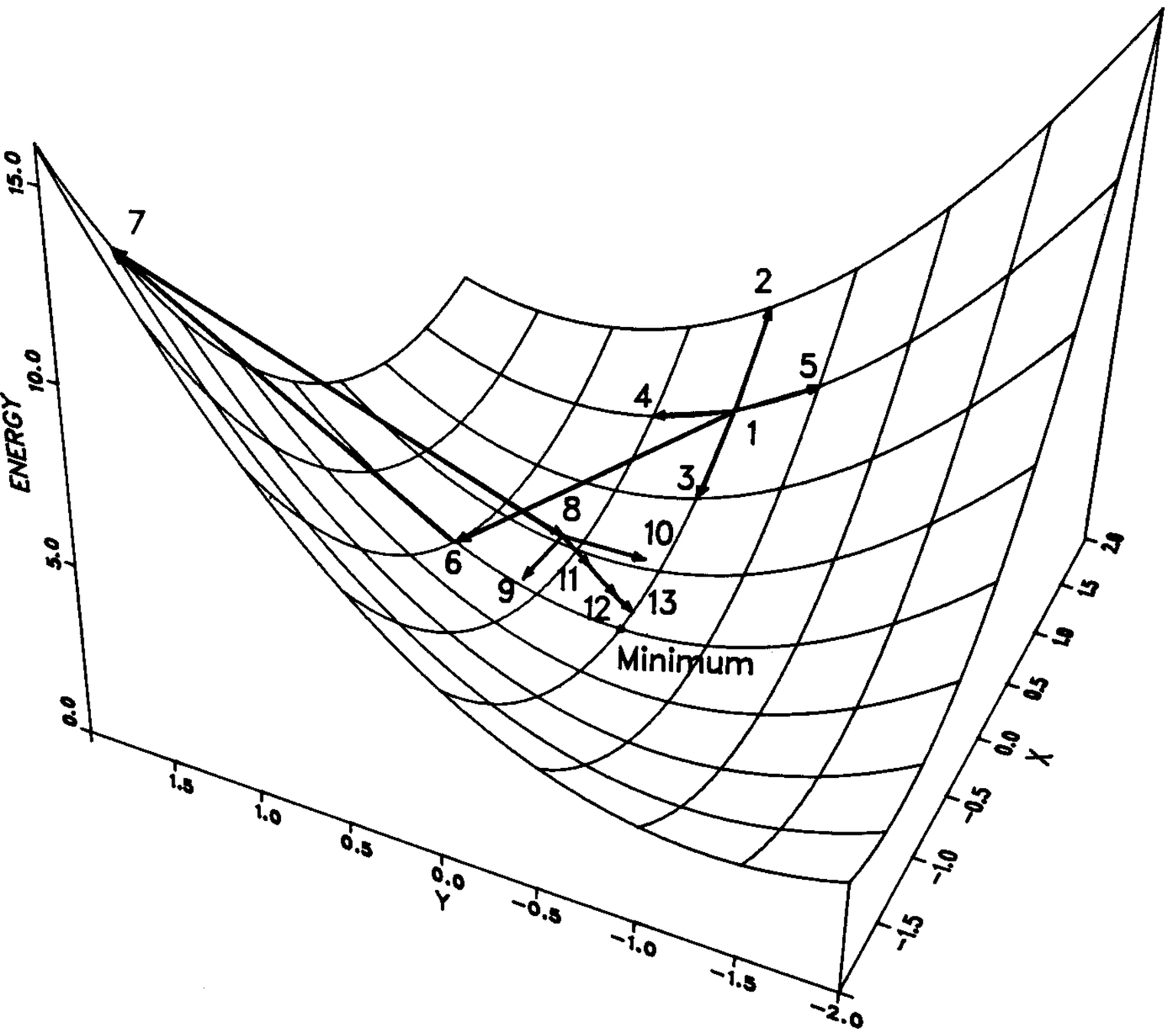

Fig. 2. Steps in finding the minimum on a quadratic surface using a modified Fletcher–Powell algorithm.

cycle, even for a quadratic surface. This is because the example has strong coupling between the two coordinates, which results in a diagonal valley rather than one parallel to the coordinate axes. A second cycle through all of the coordinates (points (8)–(13)) yields a structure closer to the true minimum. For an N-dimensional surface, $3N + 1$ energy calculations are needed for the first cycle and $3N$ for each subsequent cycle, with typically two or three cycles needed for satisfactory convergence. In general, convergence is slow and more cycles are required if the coordinates are strongly coupled and/or the eigenvalues of the second derivative matrix are very different, e.g. a long narrow valley at an angle to the coordinate axes. The only test for the accuracy of the optimization is to carry out one more cycle through all of the coordinates to determine if any change significantly.

A second, more sophisticated, method is a variant of the Fletcher–Powell algorithm[75,76] and is shown in Fig. 2. The method is actually a derivative-based method (fixed metric, using numerical derivatives) but is rather closely related to the axial iteration method. The following steps are involved:

1. Calculate the energy at the starting geometry, $\mathbf{x}_k$, $k = 0$, and at positive and negative displacements for each of the coordinates (points (1)–(5) in Fig. 2).
2. Fit a parabola for each of the coordinates, i.e. a quadratic surface without interaction terms:

$$E(\mathbf{x}) = E_k + \sum_i [g_i^k(x_i - x_i^k) + \tfrac{1}{2}B_{ii}(x_i - x_i^k)^2] \tag{1}$$

   In effect, the gradient, $\mathbf{g}$, and diagonal elements of the Hessian, $\mathbf{B}$, are calculated numerically.
3. Find the minimum on the model surface:

$$dE/dx_i = g_i(\mathbf{x}) = g_i^k + B_{ii}(x_i - x_i^k) = 0$$
$$p_i^k = x_i - x_i^k = -g_i^k/B_{ii} \tag{2}$$

   If $\mathbf{p}_k$, the predicted change in the coordinates, is small enough, stop.
4. Calculate the energy at $\mathbf{x}_k + \mathbf{p}_k$ and $\mathbf{x}_k + 2\mathbf{p}_k$ (points (6) and (7) in Fig. 2).
5. Fit a parabola to the energy at $\mathbf{x}_k$, $\mathbf{x}_k + \mathbf{p}_k$ and $\mathbf{x}_k + 2\mathbf{p}_k$, and find the minimum, $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha\mathbf{p}_k$.
6. Calculate the energy at $\mathbf{x}_{k+1}$ (point (8) in Fig. 2).
7. Compute $\mathbf{g}_{k+1}$ by calculating the energy at displacements for each coordinate from $\mathbf{x}_{k+1}$ (points (9) and (10)).
8. Set $k = k + 1$, and go back to step 3.

For an N-dimensional surface, $2N + 4 + m(N + 3)$ steps are required, where m is the number of passes through step 7. (typically two or three). A test of the goodness of an individual structure requires $N + 1$ energy computations to evaluate the first derivative numerically. As N increases, the modified Fletcher–Powell method is significantly more efficient than the axial iteration

algorithm, especially for strongly coupled coordinates where several optimization cycles may be required. As can be seen from Fig. 2, this algorithm is also not exact in one or two cycles for a general quadratic surface, because the off-diagonal terms in the Hessian are neglected.

Improvements that include the explicit calculation of the off-diagonal elements of the Hessian, B_{ij} $(i \neq j)$, have been discussed by several authors.[77,78] Because the B_{ij} are difficult to calculate accurately for a non-quadratic surface, and because there are $N(N-1)/2$ elements B_{ij}, considerable care is needed to balance the number of function evaluations with the numerical accuracy of the B_{ij}, in order to obtain a more efficient algorithm for minimization.

B. Gradient Algorithms

Quasi-Newton, variable metric, conjugate gradient,[79] Fletcher–Powell,[75] Davidon–Fletcher–Powell,[75] Murtagh–Sargent,[80] Broyden–Fletcher–Goldfarb–Shanno,[81] and optimally conditioned[82] are a few of the optimization algorithms that use the first derivative.[70] If analytical derivatives are available, these can be significantly more efficient and can have better convergence properties than the function-only algorithms. If gradients must be calculated numerically, the overall efficiency may not be better than the function-only algorithms discussed above.

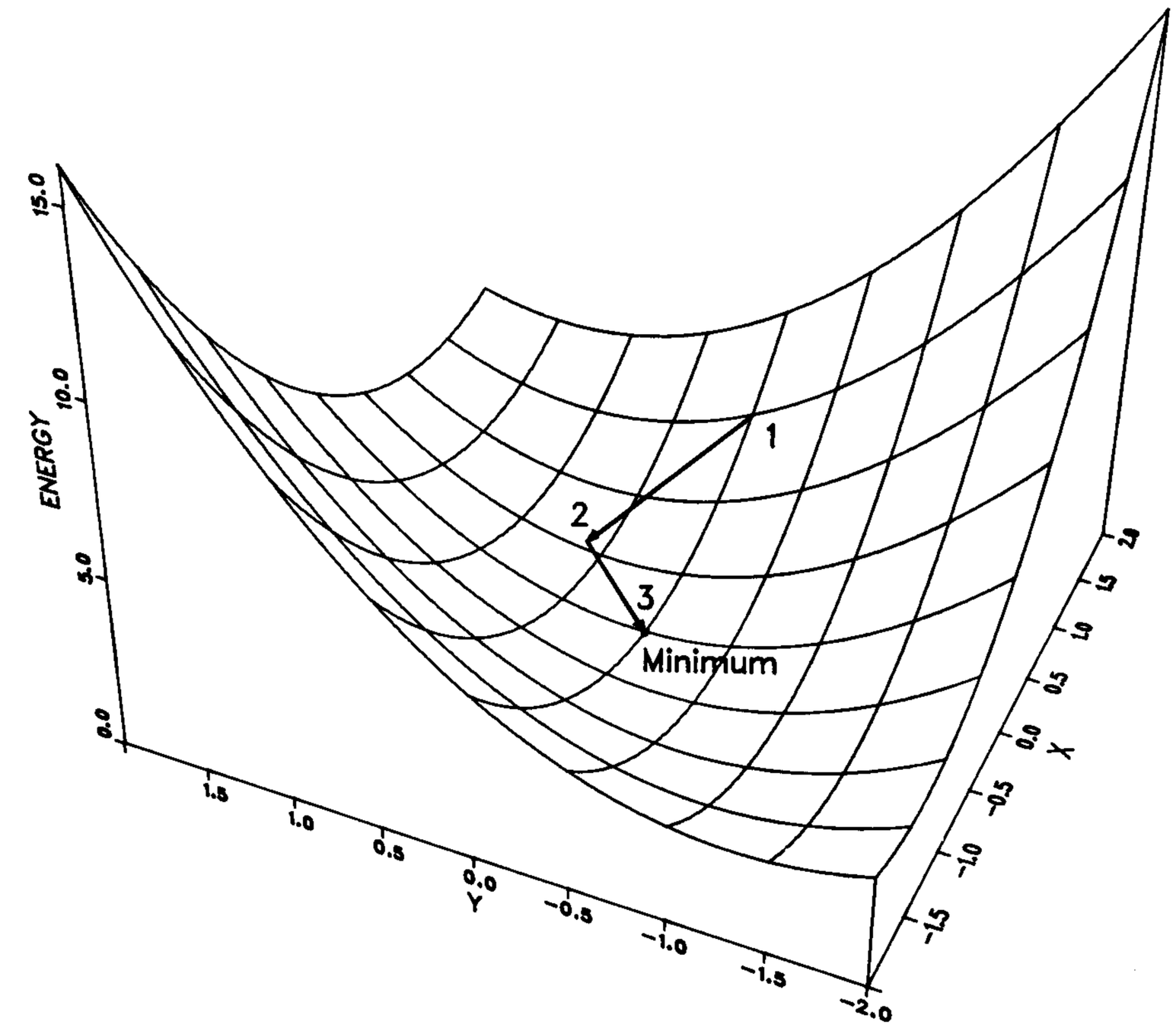

Fig. 3. Steps in finding the minimum on a quadratic surface using a quasi-Newton algorithm with accurate line searches.

The gradient-type optimization algorithms approximate the energy surface at step k by a quadratic expression in terms of the position, $\mathbf{x}_k$, the computed energy, E_k, the computed gradient, $\mathbf{g}_k$, and the approximate Hessian, $\mathbf{B}_k$:

$$E(\mathbf{x}) = E_k + \mathbf{g}_k^{\mathrm{T}} \cdot (\mathbf{x} - \mathbf{x}_k) + \tfrac{1}{2}(\mathbf{x} - \mathbf{x}_k)^{\mathrm{T}} \mathbf{B}_k (\mathbf{x} - \mathbf{x}_k) \tag{3}$$

The initial estimate of the Hessian, $\mathbf{B}_0$ (or its inverse, $\mathbf{H}_0 = \mathbf{B}_0^{-1}$), is improved as the optimization proceeds, and for most methods approaches the true Hessian (or its inverse) if the surface is quadratic. One-dimensional minimization may also be required along each new search direction. A typical search for a two-dimensional quadratic surface is shown in Fig. 3; actual optimization may be in 10–20 dimensions. An optimization can be divided into the following steps:

1. Start with the geometry $\mathbf{x}_k$, $k = 0$, and obtain an estimate of the Hessian, $\mathbf{B}$, or its inverse, $\mathbf{H}$ (either a unit matrix, an empirical estimate, or other, see below).
2. For $\mathbf{x}_k$ calculate the energy, E_k, and gradient, $\mathbf{g}_k = \mathrm{d}E/\mathrm{d}\mathbf{x}$.
3. Update the Hessian (or inverse Hessian) so that the model surface fits the current energy and gradient as well as those from previous steps (omit for the first point).
4. Find the minimum on the model surface using the gradient and the updated Hessian (note that this requires $\mathbf{H}$, the inverse of the Hessian):

$$\mathrm{d}E/\mathrm{d}\mathbf{x} = \mathbf{g}(\mathbf{x}) = \mathbf{g}_k + \mathbf{B}_k(\mathbf{x} - \mathbf{x}_k) = 0$$
$$\mathbf{p}_k = \mathbf{x} - \mathbf{x}_k = -\mathbf{B}_k^{-1}\mathbf{g}_k = -\mathbf{H}_k\mathbf{g}_k \tag{4}$$

   If the gradient, $\mathbf{g}_k$, is small enough and/or the predicted change in the geometry, $\mathbf{p}_k$, is small enough, stop.
5. Carry out a minimization in the direction of the predicted displacement, i.e. minimize $E(\mathbf{x}_k + \alpha\mathbf{p}_k)$ with respect to α (not used in some methods, i.e. $\alpha = 1$ always).
6. Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha\mathbf{p}_k$, $k = k + 1$, and return to step 2.

This sequence is illustrated for a quadratic surface by points (1)–(3) in Fig. 3 (accurate line search assumed). For an N-dimensional surface, only one gradient calculation is needed to test whether a point is a minimum. For a quadratic surface, gradient methods require $N + 1$ steps or less, and are guaranteed to reach the minimum, regardless of the starting geometry or the magnitude of the coupling between the coordinates. Accurate line searches

may be required to prove convergence for some methods. Even for the non-quadratic surfaces encountered in typical geometry optimizations, convergence is usually achieved in about $N+1$ steps, and often fewer since good estimates of the geometry and the Hessian can be made. Because an energy plus gradient calculation takes about twice as long as an energy calculation alone, each step in a gradient algorithm is more costly than in an energy-only algorithm. However, this is more than compensated by the reduced number of steps and the more stable convergence properties of the gradient optimizations. In practice, as much as an order-of-magnitude increase in efficiency can be obtained with gradient optimization methods compared to energy-only algorithms.

Aside from the nature of the function and the starting coordinates, the overall cost and convergence rate for minimizing a non-quadratic function depend on the updating scheme for the Hessian (or its inverse), the accuracy of the line searches and the initial estimate of the Hessian (or inverse). These topics are considered in the next two subsections. In the numerical analysis literature, much attention has been devoted to the problem of updating the Hessian and the linear searches.[70] For geometry optimization, good initial estimates of the Hessian can often be obtained from general concepts in chemical bonding or from lower levels of theory. An accurate initial estimate of the second derivative matrix can improve the rate of convergence significantly, but will not affect the final, optimized geometry, since the latter depends only on the gradient (where it goes to zero) and not the Hessian.

1. Updating Formulas and Line Searches

The simplest gradient method is the steepest-descent algorithm.[70] The Hessian is taken as the unit matrix (or a constant times the unit matrix) and is not updated. Thus the search is along $\mathbf{p}_k = -\mathbf{g}_k$, i.e. the direction in which the function decreases most rapidly. An accurate linear search is required at each step to achieve convergence. This method reduces the function value quite rapidly at first, but final convergence is slow. Closely related to this algorithm is the fixed metric method[70] in which the Hessian is a more general non-diagonal matrix that is not updated.

At the other end of the spectrum of gradient algorithms is Newton's method.[70] The Hessian is calculated directly at each step (hence it is a second derivative method, and is discussed in Section 4.c below). The most frequently used gradient algorithms fall between the extremes of the steepest-descent/fixed metric methods and Newton's method, and are termed variable metric or quasi-Newton methods. These methods avoid the direct calculation of the Hessian; instead they start with an approximate Hessian and improve it using the gradient information gathered during the course of the optimization. Only the main features are summarized here; a more detailed discussion can be

found in books dealing with optimization[70] and in the original literature.

The conjugate gradient (CG) algorithm[79] is one of the older methods and, strictly speaking, is not a quasi-Newton method. However, it is the method of choice for very large problems where the storage of the Hessian is not practicable. In the Fletcher–Reeves approach the search direction is given by

$$\mathbf{p}_k = -\mathbf{g}_k + \mathbf{p}_{k-1}(\mathbf{g}_k^{\mathrm{T}}\cdot\mathbf{g}_k)/(\mathbf{g}_{k-1}^{\mathrm{T}}\cdot\mathbf{g}_{k-1}) \tag{5}$$

To relate the conjugate gradient algorithm to the quasi-Newton methods this formula can be expressed[74] as an updating scheme for the inverse Hessian:

$$\mathbf{H}_k = \mathbf{H}_{k-1} - \frac{\mathbf{H}_{k-1}\Delta\mathbf{g}_k\Delta\mathbf{g}_k^{\mathrm{T}}\mathbf{H}_{k-1}}{\Delta\mathbf{g}_k^{\mathrm{T}}\mathbf{H}_{k-1}\Delta\mathbf{g}_k} \qquad \Delta\mathbf{g}_k = \mathbf{g}_k - \mathbf{g}_{k-1} \quad \text{and} \quad \mathbf{H}_0 = \mathbf{I} \tag{6}$$

If the Hessian can be kept in memory, the quasi-Newton methods provide better convergence to the minimum. Some of the more frequently used schemes for updating the Hessian are as follows.

1. The Davidon–Fletcher–Powell (DFP) algorithm:[75]

$$\mathbf{H}_k = \mathbf{H}_{k-1} + \frac{\Delta\mathbf{x}_k\Delta\mathbf{x}_k^{\mathrm{T}}}{\Delta\mathbf{x}_k^{\mathrm{T}}\cdot\Delta\mathbf{g}_k} - \frac{\mathbf{H}_{k-1}\Delta\mathbf{g}_k\Delta\mathbf{g}_k^{\mathrm{T}}\mathbf{H}_{k-1}}{\Delta\mathbf{g}_k^{\mathrm{T}}\mathbf{H}_{k-1}\Delta\mathbf{g}_k} \qquad \Delta\mathbf{x}_k = \mathbf{x}_k - \mathbf{x}_{k-1} \tag{7}$$

2. The Murtagh–Sargent (MS) algorithm:[80]

$$\mathbf{H}_k = \mathbf{H}_{k-1} + \frac{[\Delta\mathbf{x}_k - \mathbf{H}_{k-1}\Delta\mathbf{g}_k][\Delta\mathbf{x}_k - \mathbf{H}_{k-1}\Delta\mathbf{g}_k]^{\mathrm{T}}}{[\Delta\mathbf{x}_k - \mathbf{H}_{k-1}\Delta\mathbf{g}_k]^{\mathrm{T}}\cdot\Delta\mathbf{g}_k} \tag{8}$$

3. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm:[81]

$$\mathbf{H}_k = \left[\mathbf{I} - \frac{\Delta\mathbf{x}_k\Delta\mathbf{g}_k^{\mathrm{T}}}{\Delta\mathbf{x}_k^{\mathrm{T}}\cdot\Delta\mathbf{g}_k}\right]\mathbf{H}_{k-1}\left[\mathbf{I} - \frac{\Delta\mathbf{x}_k\Delta\mathbf{g}_k^{\mathrm{T}}}{\Delta\mathbf{x}_k^{\mathrm{T}}\cdot\Delta\mathbf{g}_k}\right]^{\mathrm{T}} + \frac{\Delta\mathbf{x}_k\Delta\mathbf{x}_k^{\mathrm{T}}}{\Delta\mathbf{x}_k^{\mathrm{T}}\cdot\Delta\mathbf{g}_k} \tag{9}$$

The latter can be written in a more general form, which defines the Broyden family of algorithms:[83]

$$\mathbf{H}_k = \mathbf{H}_{k-1} + \frac{\Delta\mathbf{x}_k\Delta\mathbf{x}_k^{\mathrm{T}}}{\Delta\mathbf{x}_k^{\mathrm{T}}\cdot\Delta\mathbf{g}_k} - \frac{\mathbf{H}_{k-1}\Delta\mathbf{g}_k\Delta\mathbf{g}_k^{\mathrm{T}}\mathbf{H}_{k-1}}{\Delta\mathbf{g}_k^{\mathrm{T}}\mathbf{H}_{k-1}\Delta\mathbf{g}_k} + \pi_k\Delta\mathbf{g}_k^{\mathrm{T}}\mathbf{H}_{k-1}\Delta\mathbf{g}_k\mathbf{w}_k\mathbf{w}_k^{\mathrm{T}} \tag{10}$$

$$\mathbf{w}_k = \Delta\mathbf{x}_k/(\Delta\mathbf{x}_k^{\mathrm{T}}\cdot\Delta\mathbf{g}_k) - \mathbf{H}_{k-1}\Delta\mathbf{g}_k/(\Delta\mathbf{g}_k^{\mathrm{T}}\mathbf{H}_{k-1}\Delta\mathbf{g}_k)$$

Note that $\pi_k = 0$ yields the DFP method and $\pi_k = 1$ the BFGS method. The optimally conditioned (OC) method[82] chooses π_k to minimize the condition number of the Hessian (the condition number is the ratio of the largest to the smallest eigenvalue), thereby improving the behavior of the optimization. The CG, MS and DFP methods are also special cases of the Huang family of algorithms.[84] Equations (7)–(10) can also be used to update the Hessian, **B**, rather than its inverse, **H**, provided that Δx and Δg are interchanged when **H** is replaced by **B**.[74]

All of the updating formulas have been devised to assure that the Hessian remains symmetric and positive-definite. For non-quadratic functions, the Hessian may have to be reset, if the Hessian is no longer positive-definite, or if Eq. (4) does not yield a search direction leading to a lower value of the function. Convergence can be proven for quadratic functions in $N+1$ steps; for non-quadratic functions, more steps may be required, but convergence to a stationary point is guaranteed. In numerical tests of the various quasi-Newton methods, the BFGS and OC algorithms appear to be the methods of choice.[70]

For some algorithms, such as conjugate gradient, accurate line searches are necessary. However, for most quasi-Newton methods the line search does not have to be very accurate. Reduction of the directed gradient in the line search by a factor of 0.5 or 0.1 is often sufficient.[70] Frequently this is satisfied by fixing $\alpha = 1$ (i.e no line search). A few more quasi-Newton steps may be needed, but the number of function evaluations in the line search is greatly reduced. The result is a considerable savings in the total number of function and gradient evaluations needed to reach the minimum. On the other hand, if computation of the gradient is expensive (e.g. numerical gradient for a function of many variables), there is some advantage to accurate line searches using only the function value to reduce the number of costly gradient evaluations.

The line searches can be improved with little additional cost by fitting a function to the energy and gradient at $\alpha = 0$ (i.e. $\mathbf{x}_k$) and $\alpha = 1$, and interpolating the energy and gradient at the minimum. A cubic polynomial,[85] a quartic polynomial constrained to have only one minimum[86] and a conic[87] have been used for this purpose. Alternatively, the one-dimensional search need not be along a linear path. McKelvey and Hamilton[88] have suggested searching along a path $\mathbf{p}_k(s)$ given by

$$\frac{d\mathbf{p}_k(s)}{ds} = \mathbf{g}_k + \mathbf{B}_k\mathbf{p}_k(s) \qquad \mathbf{p}_k(s) = \left[\int_0^s \exp(\mathbf{B}_k\sigma)\,d\sigma\right]\mathbf{g}_k \tag{11}$$

At $s \to 0$, this path follows the steepest-descent direction; at large negative s, the limit is Newton's method.

Two other gradient methods that have been used for geometry optimization do not fall into the schemes outlined above. The algorithm proposed by Császár and Pulay[89] uses a fixed Hessian with a k-dimensional search at the kth iteration. The best linear combination of the current and previous geometries is chosen

$$\mathbf{x}_{k+1} = \sum_{i=0}^{k} c_i\mathbf{x}_i \qquad \sum_{i=0}^{k} c_i = 1 \tag{12}$$

such that the predicted displacement

$$\mathbf{p}_k = -\mathbf{B}^{-1}\sum_{i=0}^{k} c_i\mathbf{g}_i \tag{13}$$

is a minimum in a least-squares sense (i.e. $\mathbf{p}_k^T\cdot\mathbf{p}_k$ is a minimum with respect to c_i). Convergence is guaranteed for any positive-definite starting Hessian and appears to be better than some of the quasi-Newton methods in selected geometry optimization problems.[89]

A method devised by Schlegel[86] has seen widespread use as a part of the GAUSSIAN system of MO programs.[30] The Hessian in the full space is updated in the smaller space spanned by the current and previous steps. The available gradients are projected into the small space and elements of the Hessian in the small space are computed according to

$$b_{ij} = \frac{(\mathbf{g}_i - \mathbf{g}_k)^T\cdot\mathbf{r}_j - \sum_{m=i+1}^{k-1} b_{mj}[(\mathbf{x}_i - \mathbf{x}_k)^T\cdot\mathbf{r}_m]}{(\mathbf{x}_i - \mathbf{x}_k)^T\cdot\mathbf{r}_i}$$

$$j \leqslant i, \quad i = k-1, k-2, \ldots \tag{14}$$

where the unit vectors for the small space are given by

$$\tilde{\mathbf{r}}_i = (\mathbf{x}_i - \mathbf{x}_k) - \sum_{m=i+1}^{k-1} \mathbf{r}_m[(\mathbf{x}_i - \mathbf{x}_k)^T\cdot\mathbf{r}_m] \tag{15}$$

$$\mathbf{r}_i = \tilde{\mathbf{r}}_i/|\tilde{\mathbf{r}}_i| \qquad i = k-1, k-2, \ldots$$

The correction to the Hessian in the full space is

$$\mathbf{B}_k = \mathbf{B}_{k-1} + \sum_{i\leqslant j}[b_{ij} - \mathbf{r}_i^T\mathbf{B}_{k-1}\mathbf{r}_j][\mathbf{r}_i\mathbf{r}_j^T - (1-\delta_{ij})\mathbf{r}_j\mathbf{r}_i^T] \tag{16}$$

At step k, this results in a rank k update to the Hessian, as opposed to the rank 1 and 2 formulas used in the quasi-Newton methods, leading to improved convergence properties. The line search is avoided by using a constrained quartic polynomial to estimate the position of the minimum. The energy and gradient at the line search minimum are obtained by interpolation rather than by recalculation.

2. *Estimating the Hessian Matrix*

An initial estimate for the Hessian (or its inverse) is required for most gradient methods. The estimate need not be very accurate (e.g. a unit matrix), since the Hessian is updated during the search for the minimum, and gradually approaches the correct second derivative matrix. However, the overall efficiency of the optimization and the rate of convergence to the equilibrium geometry can be improved considerably by a good initial estimate of the Hessian (and the starting geometry). Unlike the general functional minimization problem, the task of geometry optimization is simplified by a hierarchy of theoretical methods that give increasingly more accurate descriptions of the energy surface. A lower level of theory can be used to provide a suitable,

inexpensive initial estimate of the Hessian (and the starting geometry) for optimization at a higher level of theory. Several options of increasing complexity (and cost) can be considered.

1. The unit matrix. Although this is an unbiased choice, all the useful structural information about the molecule is discarded, i.e. the nature of the atoms, the bonds between them, etc. Flexible coordinates (i.e. torsion and ring deformation) are not distinguished from stiff modes (i.e. bond stretching) and all coupling between coordinates is ignored. Such information must be accumulated during the course of the optimization at the expense of additional optimization steps. This is particularly detrimental for cyclic molecules whose coordinates are inherently strongly coupled.
2. An empirical guess. For bond stretches, the diagonal elements of the second derivative matrix can be obtained by relating bond length to stretching force constant using Badger's rule;[90] for angle bends and torsions a constant is sufficient. For cyclic molecules, a simple valence force field in redundant internal coordinates, transformed to the non-redundant coordinates used for the optimization, can provide the most important off-diagonal elements.[22] The empirical force fields used in spectroscopy and molecular mechanics are only of limited value for estimating the Hessian, since the parametrizations are usually too specialized to cover the range of molecules encountered in *ab initio* geometry optimizations.
3. A Hessian computed by semi-empirical MO methods. Some empirical adjustment of the second derivatives is usually necessary, since the semi-empirical methods tend to overestimate some terms and underestimate others.
4. Numerical calculation of a few of the more important elements of the second derivate matrix from energy or energy and gradient calculations at small displacements from the initial geometry.
5. An analytically computed second derivative matrix from a lower level of *ab initio* calculation (e.g. from a smaller basis calculation or, in the case of an optimization including correlation energy, from an SCF calculation). Similarly, the approximate Hessian from an optimization at a lower level can provide a good initial estimate of the Hessian.
6. The full second derivative matrix calculated analytically or numerically. This is the most costly and most accurate option, but may be the method of choice for difficult cases.

C. Second Derivatives

If the second derivatives are available, the minimum for a quadratic surface can be found in one step by Newton's method:[70]

$$\mathbf{x}_{\min} = \mathbf{x}_k - \mathbf{B}_k^{-1}\mathbf{g}_k = \mathbf{x}_k - \mathbf{H}_k\mathbf{g}_k \tag{17}$$

For non-quadratic surfaces, more than one step will be needed, but often only the gradient and not the second derivative matrix has to be recalculated on subsequent steps. If necessary, the second derivative matrix can be recalculated every few steps, or can be updated as in the gradient optimization algorithms. The latter is identical to using the exact second derivative matrix as the starting estimate of the Hessian in the gradient method. If second derivatives are not available (e.g. at the MP2 level), those calculated at a lower level of theory (e.g. at the SCF level) may be sufficient to assure rapid convergence to the optimized geometry. For a well behaved minimization, the Hessian must be positive-definite and not be ill-conditioned (i.e. not have a very wide spread in eigenvalues). If these conditions are not fulfilled, the problem can be remedied by adding a constant times the unit matrix to the Hessian (cf. level shifting to solve SCF convergence problems) or by adding a constant to the offending eigenvalue(s).

One of the main drawbacks to optimization methods based on second derivatives is that analytical second derivatives are available only for a few levels of theory, as discussed above. Furthermore, the computer time required for analytical second derivatives can be 5–10 to N times as long as a gradient calculation. Since algorithms using first derivatives often converge in about N steps, there may be no significant cost advantage to using second derivative based algorithms. However, analytical second derivatives may be more cost effective when there are convergence difficulties in the optimization, such as for shallow wells, strongly coupled coordinates and narrow curved valleys, or when second derivatives are only a few times as expensive as the energy and gradient evaluations (e.g. MCSCF calculations).

V. LOCATING TRANSITION STRUCTURES

Both minima and saddle points are stationary points characterized by a zero gradient. However, unlike a minimum, a first-order saddle point must be a maximum along one (and only one) direction. In general, this direction is not known in advance and must be determined during the course of the optimization. Numerous algorithms have been proposed to deal with the problem of locating transition structures.[91–93] In this section a few of the more widely used methods are surveyed. The algebraic details can be found in the original literature.

A. Surface Fitting

The simplest approach to locating a transition structure is to fit a suitable analytical expression to a set of computed energies near the transition structure. The analytical derivatives of the fitted function can then be used to

locate the transition structure on the fitted surface, without carrying out additional and expensive molecular-orbital calculations. This method is feasible for transition structures in small molecules.[78] However, surface fitting has several disadvantages. There are no standard, universally applicable methods for fitting multi-dimensional non-quadratic surfaces; each family of reactions is a special case if a large region of the surface is needed. A large number of energy (or energy and gradient) calculations are required to obtain an acceptable fit for the model energy surface in more than two or three variables. Finally, the fitted surface may not be sufficiently good in the region of the barrier to yield an accurate estimate of the geometry of the transition state. However, there are circumstances when the construction of an analytical energy surface may be necessary. An accurate fit to a large portion of the energy surface is needed for the study of reaction dynamics by classical or semiclassical trajectory calculations.[94] Smaller regions near the transition structure are required for variational transition state theory[15] and reaction path Hamiltonian methods.[16]

B. Linear and Quadratic Synchronous Transit

The linear synchronous transit (LST) and quadratic synchronous transit (QST) methods proposed by Halgren and Lipscomb[95] simplify the problem of finding the transition state by making some assumptions about the reaction path. The LST approach assumes the path is a straight line connecting reactants and products. The LST estimate of the transition state, structure 1 in Fig. 4, is the energy maximum along the straight-line path. The energy of 1 is an upper bound to the true saddle point energy, often a rather poor bound if the actual path is curved. The estimate of the transition structure can be improved by minimizing the energy with respect to all of the coordinates perpendicular to the linear reaction path. The resulting point, 2, is a lower bound to the true saddle point. The reaction path can now be approximated by a parabolic path between reactants and products, i.e. a quadratic synchronous transit (QST) path.[95] The maximum on the quadratic synchronous path, 3, yields a much better estimate of the energy and position of the transition state. Under less favorable circumstances (examples are discussed elsewhere[92,93]), the process of minimizing perpendicular to the path and finding a maximum along the quadratic transit path must be repeated a number of times before satisfactory convergence is obtained. If the actual reaction path is very curved, the QST may also be a poor approximation to the true path between reactants and products, and the method may not converge to a satisfactory structure. In part this can be remedied if structures closer to the transition state are used as end points of the LST and QST paths.[95] A number of algorithms have been devised[93,96–100] that alternate between maximizing along a path of predetermined, constrained form and minimizing perpendicular to the path. Methods

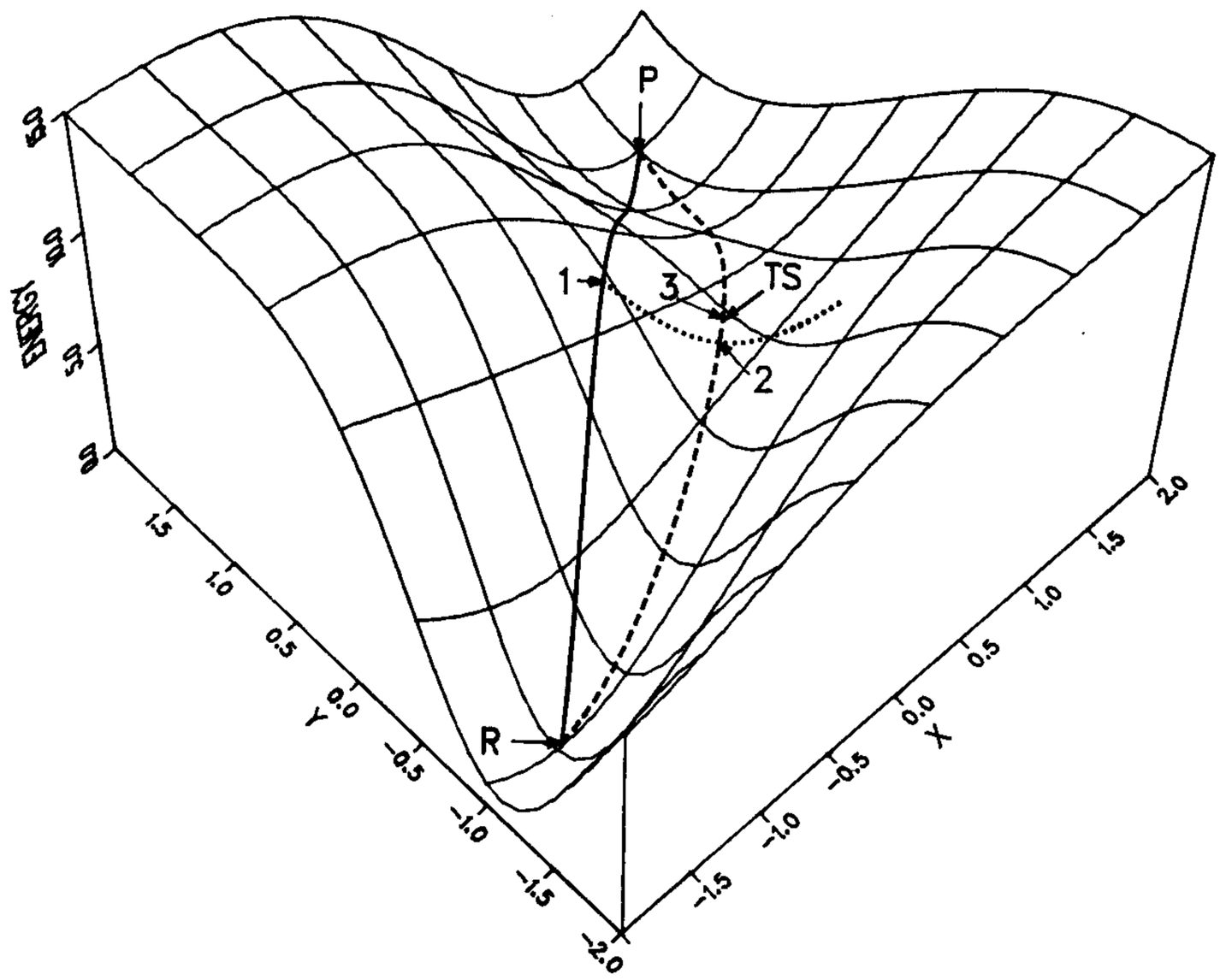

Fig. 4. The linear synchronous transit (LST) and quadratic synchronous transit (QST) methods for finding transition structures: R, reactants; P, products; TS, true transition structure; 1, maximum on LST path (full curve); 2, minimum perpendicular to LST path; 3, maximum on QST path (broken curve). The model surface is constructed from two Gaussians.

using gradient algorithms for the minimization step are considerably more efficient than non-gradient methods.

C. Coordinate Driving

For many reactions, the transformation from reactants to products is dominated by a change in one coordinate. A series of points along the reaction path can be obtained by stepping along the dominant coordinate and minimizing the energy with respect to the remaining $N-1$ coordinates (Fig. 5). In favorable cases,[101] the path generated passes through the transition structure. Provided the step size is sufficiently small near the saddle point, the energy maximum along the path can be a good approximation to the transition structure. The coordinate driving method can be considered an extension of the QST method because the path is not constrained to be quadratic. However, coordinate driving is rather costly, since an optimization of $N-1$ coordinates is required for each step taken. Furthermore, if the path is too curved and becomes dominated by one of the coordinates being

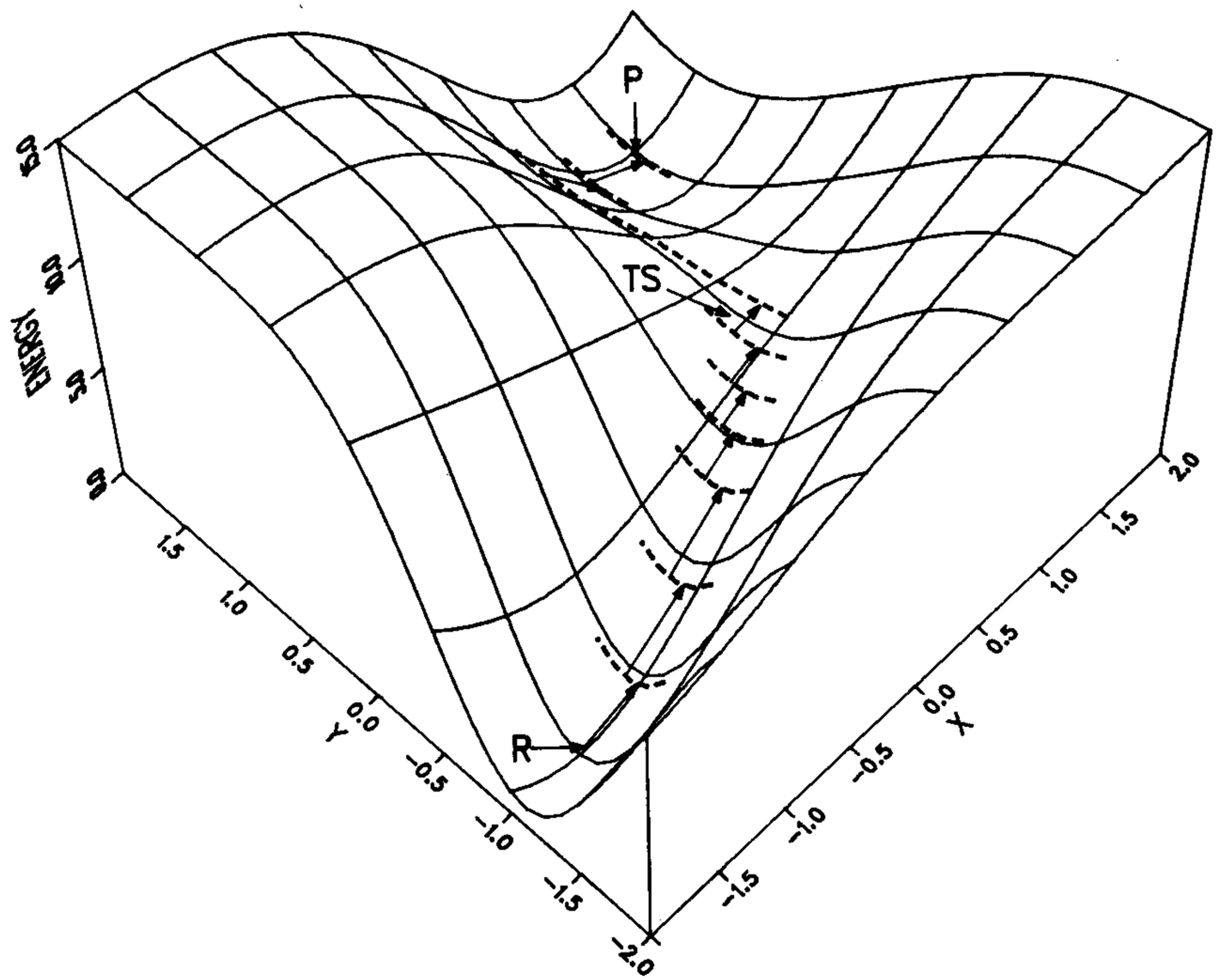

Fig. 5. The coordinate driving method for locating transition structures. A series of points is generated by stepping one coordinate (X) and optimizing the other (Y): R, reactants; P, products; TS, transition structure.

minimized, the method can encounter serious difficulties.[102] If the transition structure is reached, as in Fig. 5, there can be an abrupt change in the geometry just beyond the transition structure, because the minimization is now being carried out along the reaction path rather than perpendicular to it. In other cases (Fig. 6), coordinate driving may miss the transition structure entirely, if the saddle point is off to the side of the main valley. Sometimes (but not always) these problems can be overcome by starting from the products instead of the reactants, and/or using a different driving coordinate.

D. Hill Climbing or Walking Up Valleys

The difficulties encountered by the coordinate driving method when the reaction path is strongly curved can be avoided with a simple change in strategy that does not add appreciably to the computational effort. Instead of a step along a specific coordinate, a step of fixed length is taken in the direction that gives the easiest (least steep) path up the valley,[103–108] as shown in Fig. 7. Near the saddle point the step size can be reduced to locate the stationary point more accurately. Care must be taken that the step length does not

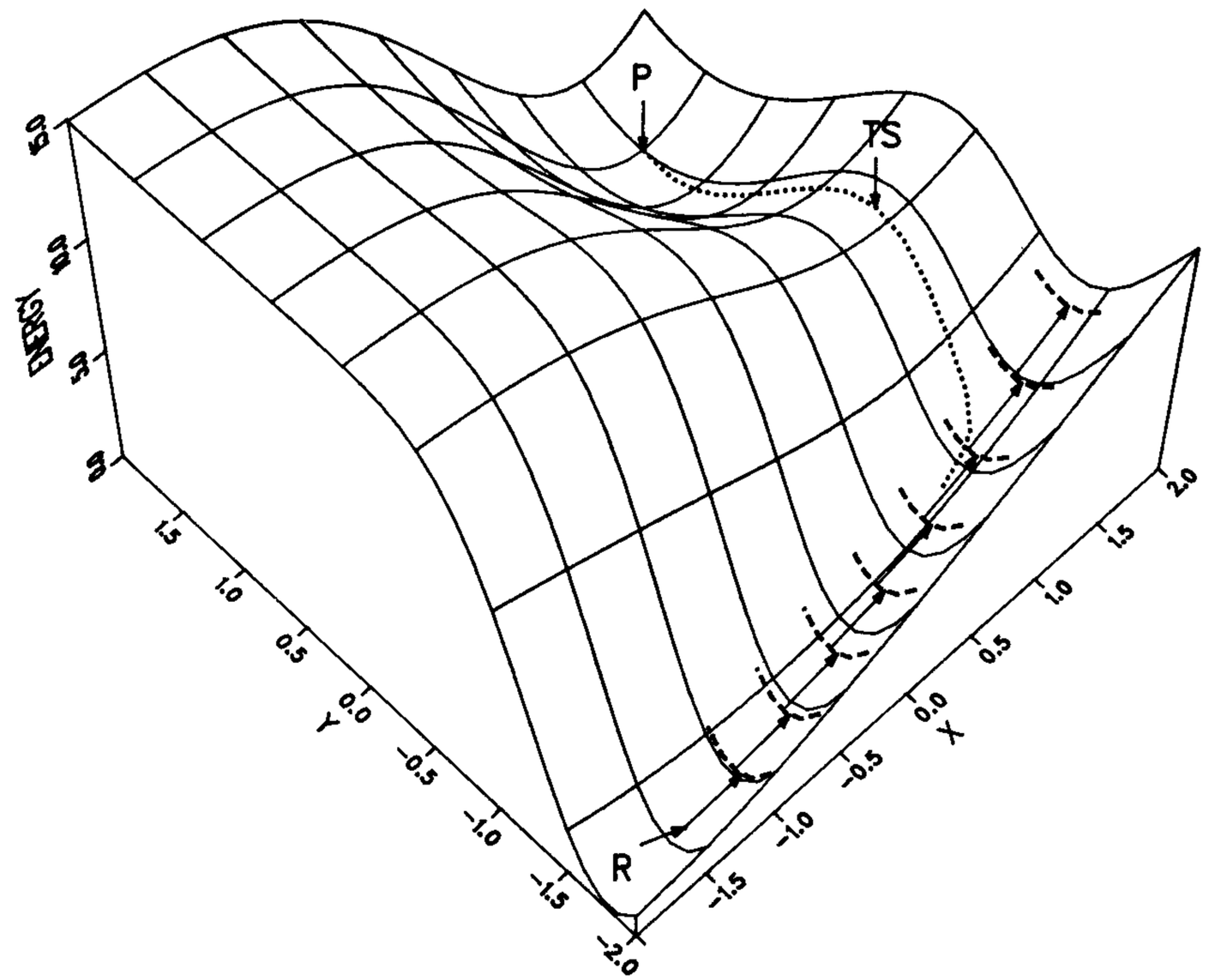

Fig. 6. Failure of coordinate driving method due to a long valley not leading to the transition structure. The dotted curve is the steepest-descent path from the saddle point.

become smaller than the radius of curvature of the contour lines crossing the reaction path (otherwise there will be no minimum-energy uphill step). Alternatively, the step size can be changed similar to a binary search to find the transition state in fewer steps.[103] Like coordinate driving, this algorithm is costly since numerous steps may be needed to approach the transition structure, and each step requires the optimization of $N-1$ variables. Some applications of this method rely only on the energy[103] or the energy and gradients.[104,105] Other implementations[106–108] calculate the second derivative matrix at each step or every few steps in order to define the uphill path. The direction of shallowest ascent is given by the eigenvector of the Hessian with the smallest eigenvalue. The frequent calculation of the second derivative matrix makes the latter implementation very costly and currently limits it to SCF energy surfaces. In many ways, walking up a valley to the transition structure is the converse of finding the path of steepest descent from the transition structure to the reactants. However, unlike the steepest-descent path, which is guaranteed to find a minimum, a shallowest-ascent path need not find the desired transition state. Like the coordinate driving method, walking up valleys can bypass the transition structure for some surfaces, such as the one shown in Fig. 6.

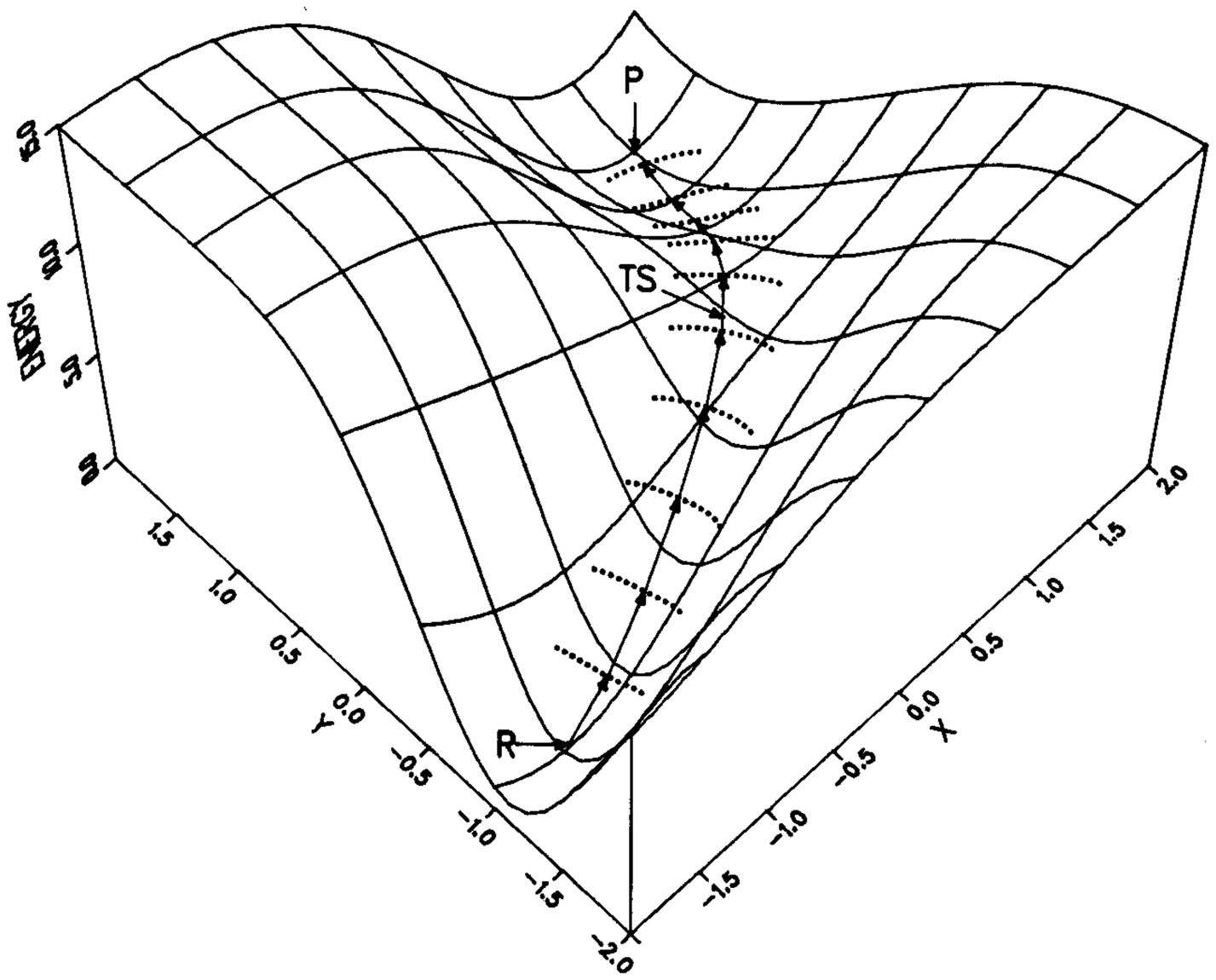

Fig. 7. Hill climbing or walking up valleys to locate transition structures. A series of points is generated by taking a fixed length step and finding the shallowest path uphill: R, reactants; P, products; TS, transition structure.

E. Gradient Norm Method

The gradient at a saddle point is zero, but standard gradient minimization methods cannot be used because the Hessian is not positive-definite. However, a saddle point optimization can be changed into a minimization by recognizing that the norm of the gradient is a minimum at any stationary point (minimum, saddle point or local maximum). In the method proposed by McIver and Komornicki,[109-111] a saddle point is found by minimizing the square of the gradient norm, $|\mathbf{g}_k^T \cdot \mathbf{g}_k|$. This is a non-linear least-squares problem, and is usually solved using Powell's algorithm[112] or a similar method. The optimization converges directly to the transition structure, provided that the starting point is near the saddle point. At the saddle point the gradient norm must be zero, not just a minimum. As shown in Fig. 8, the gradient norm can be a minimum at a shoulder, in addition to being zero at a stationary point. Care must be taken to avoid converging on such areas of the reaction surface. Figure 9 illustrates the gradient norm squared for the surface used in Figs. 4, 5 and 7. As can be seen, a very good guess for the transition structure may be needed for the gradient norm algorithm to converge to the

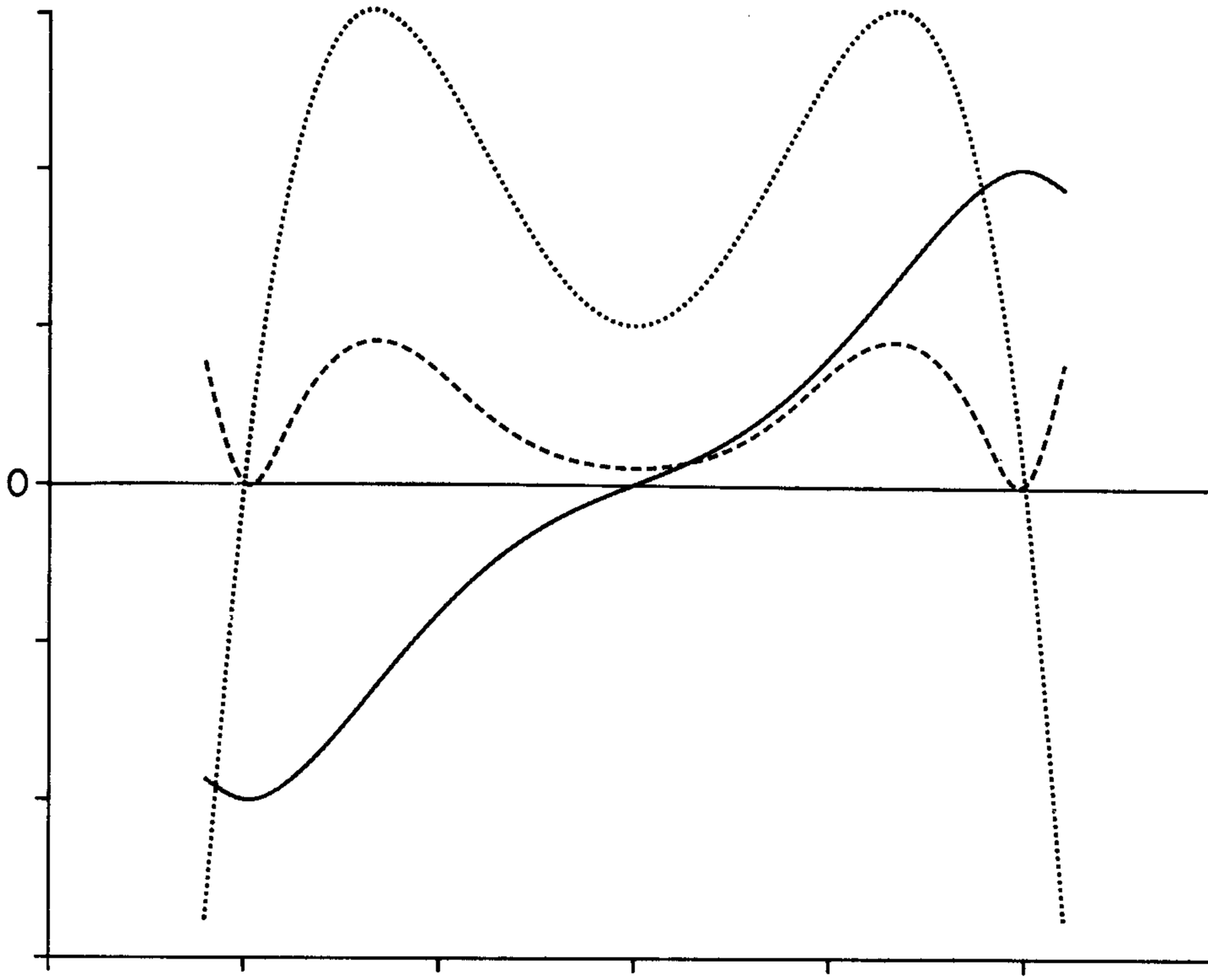

Fig. 8. A potential energy curve (full curve), its gradient (dotted curve) and its gradient norm squared (broken curve). The gradient is zero at the minimum and maximum. The gradient norm squared is a minimum not only at the potential energy minimum and maximum but also at the shoulder.

saddle point, rather than to some other feature on the surface. Furthermore, the region where the Hessian for the gradient norm is positive-definite is smaller than the quadratic region of the saddle point where the Hessian for the energy has one negative eigenvalue.

F. Gradient Algorithms

Gradient methods are very efficient for minimizations; however, to locate transition structures they must be modified or constrained to overcome the problems caused by the Hessian not being positive-definite. One approach is to partition the N-dimensional optimization into a one-dimensional space for maximization, and an $(N-1)$-dimensional space for minimization.[93,96,99,100] This partitioning in effect chooses a transition vector. The transition vector may be fixed, or may be allowed to vary in a restricted manner (e.g. according to a quadratic synchronous transit path). The search for a maximum in the

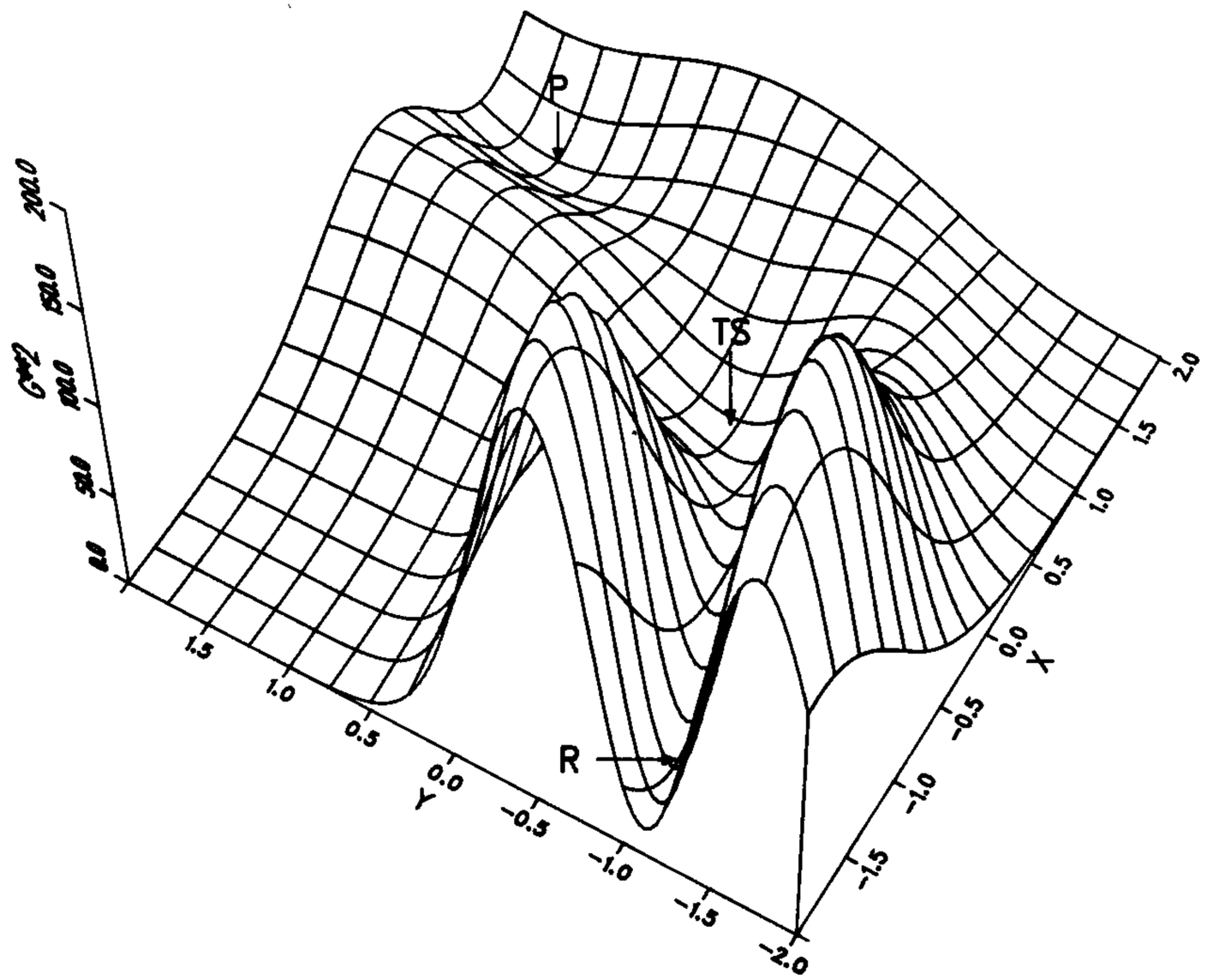

Fig. 9. The gradient norm squared for the model potential energy surface used in Figs. 4, 5 and 7. The gradient norm squared is a local minimum in many places in addition to the minima and saddle point on the potential.

one-dimensional space causes no problem. Minimization in the $(N-1)$-dimensional space can be carried out efficiently using gradient methods as discussed above. The maximization and minimization can be done simultaneously or alternately, but the final structure must be a stationary point in the full space. A drawback of this approach is that the partitioning must be chosen *a priori*. If the one-dimensional space does not contain a significant component of the true transition vector, the energy surface may not have a maximum in the one-dimensional space after several optimization steps.

An alternative approach[86] allows the transition vector to vary without constraint during the optimization. The Hessian is updated as in the minimization procedures, but must be tested to ensure that it has one and only one negative eigenvalue. If not, the offending eigenvalue is changed and the Hessian reconstructed from the eigenvectors and the modified eigenvalues. The initial Hessian must have one negative eigenvalue. An approximate Hessian can be obtained by computing a few rows of the Hessian numerically for the variables expected to dominate the transition vector, or by computing the Hessian at a lower level of theory (options 4 and 5, respectively, in Section IV.B.2).

Gradient methods for optimizing transition structures appear to have wider ranges of convergence than the gradient norm algorithm, do not have the problem of converging onto shoulders, and are considerably more efficient. However, for reliable convergence the starting geometry must be in the quadratic region of the transition state, i.e. where the Hessian has one negative eigenvalue. This may be a significant problem if the barrier is very narrow or the region of the transition state is very flat.

Gradient methods discussed above use a quadratic function (energy, gradient and approximate Hessian) to model the energy surface near the transition state. Distance-weighted interpolants[113] provide a more flexible functional form that can interpolate arbitrarily spaced points with a smooth differentiable function. For a gradient-based optimization, the Shepard interpolation functions[114] seem appropriate

$$E(\mathbf{x}) = \sum W_i(E_i + \mathbf{g}_i^{\mathrm{T}}\cdot(\mathbf{x}-\mathbf{x}_i) + \tfrac{1}{2}(\mathbf{x}-\mathbf{x}_i)^{\mathrm{T}}\mathbf{B}_k(\mathbf{x}-\mathbf{x}_i))/\sum W_i$$
$$W_i = 1/|\mathbf{x}-\mathbf{x}_i|^2 \tag{18}$$

although other related forms can be constructed. The energies and gradients are interpolated exactly with this function, even for non-quadratic surfaces. The Hessian, $\mathbf{B}_k$, is updated in a manner analogous to the gradient methods. In preliminary applications,[115] this method appears to have a wider range of convergence than the above gradient algorithms. The starting guess for the transition state geometry need not be one structure, but can be a set of structures bracketing the transition state.

G. Second Derivatives

If second derivatives of the energy are available, then Newton's method, Eq. (17), can be used to locate saddle points in a manner analogous to finding minima. However, the Hessian must be tested at each step to insure that it has one and only one negative eigenvalue. For the method to converge without difficulties, the starting structure must be in the quadratic region of the transition state. If the Hessian has the wrong number of negative eigenvalues, either at the starting structure or at any point during optimization, then the Hessian must be adjusted so that it has one negative eigenvalue. This may be done by changing the sign of the offending eigenvalue (usually the smallest one) or by adding/subtracting a constant. The adjusted Hessian is then reconstructed from the eigenvectors and the modified eigenvalues. If the Hessian is initially positive-definite and the smallest eigenvalue is made negative, the result is the same as the algorithm for walking up valleys based on the computation of second derivatives, discussed above. As already cautioned, such procedures can get lost in dead-end valleys. Although the computation of second derivatives is more expensive than gradients, a second derivative-based

approach may be the method of choice for difficult problems. Similar to minimization, it may not be necessary to recalculate the second derivatives at every step (the Hessian can be left constant, updated, or recomputed every few steps).

H. Testing Stationary Points

Except for relatively simple surfaces, it is usually necessary to verify that the stationary point found in a transition structure optimization is indeed a first-order saddle point and not a minimum or a higher-order saddle point. This can be done by calculating the full Hessian (analytically or numerically) at the stationary point and checking that there is one and only one negative eigenvalue. Note that the approximate Hessian obtained by an updating procedure in a gradient optimization on a non-quadratic surface is not sufficiently accurate to test the nature of the stationary point. Furthermore, it contains no information about displacements that lead to lower-symmetry, and possibly lower-energy, saddle points. If the full Hessian has a second negative eigenvalue, then one of these two eigenvectors is not a suitable transition vector for the reaction. A lower-energy saddle point can be obtained by displacing along this eigenvector and re-optimizing the geometry.

Once a saddle point has been found and characterized by computing the Hessian, it must still be tested to determine that it is the transition structure for the desired reaction. The transition vector may indicate clearly enough that the reaction path connects the desired reactants and products. If not, it may be necessary to follow the reaction path down from the saddle point (perhaps only for a short distance) to be sure that the path leads to the correct reactants and products.

VI. REACTION PATH FOLLOWING

At the saddle point, the intrinsic reaction coordinate (IRC) in mass-weighted Cartesian coordinates is coincident with the transition vector (i.e. the eigenvector of the Hessian with the negative eigenvalue).[13] On either side of the saddle point, the IRC in mass-weighted Cartesian coordinates is the steepest-descent path from the saddle point to the reactants and to the products.[13] However, geometry optimizations are often more convenient in internal coordinates, and the steepest-descent path in internal coordinates is not the IRC. To obtain the IRC in internal coordinates, the transformation from mass-weighted Cartesian coordinates to internal coordinates must be taken into account.[13,116] The descent direction ($-\mathbf{g}$ in mass-weighted Cartesian coordinates) is then given by $-\mathbf{Gg}$, where $\mathbf{G}$ is Wilson's G matrix (note that $G_{ij} = \delta_{ij}$ for mass-weighted Cartesian coordinates and $G_{ij} = m_i^{-1} d_{ij}$ for Cartesian coordinates, where m_i are the masses of the atoms). In the formulas below, $\mathbf{g}$ must be replaced by $\mathbf{Gg}$ if mass-weighted Cartesian coordinates are not used, and the transition vector must be replaced by the normal coordinate for the vibration with imaginary frequency.

The simplest approach for following the IRC is just to take a series of small steps in the descent direction:[117]

$$\mathbf{x}_{k+1}^{\mathrm{IRC}} - \mathbf{x}_k^{\mathrm{IRC}} = -s\mathbf{g}_k/|\mathbf{g}_k| \tag{19}$$

Unfortunately, except for very small step sizes, this path quickly departs from the IRC and tends to oscillate about it. Equation (19) is just a first

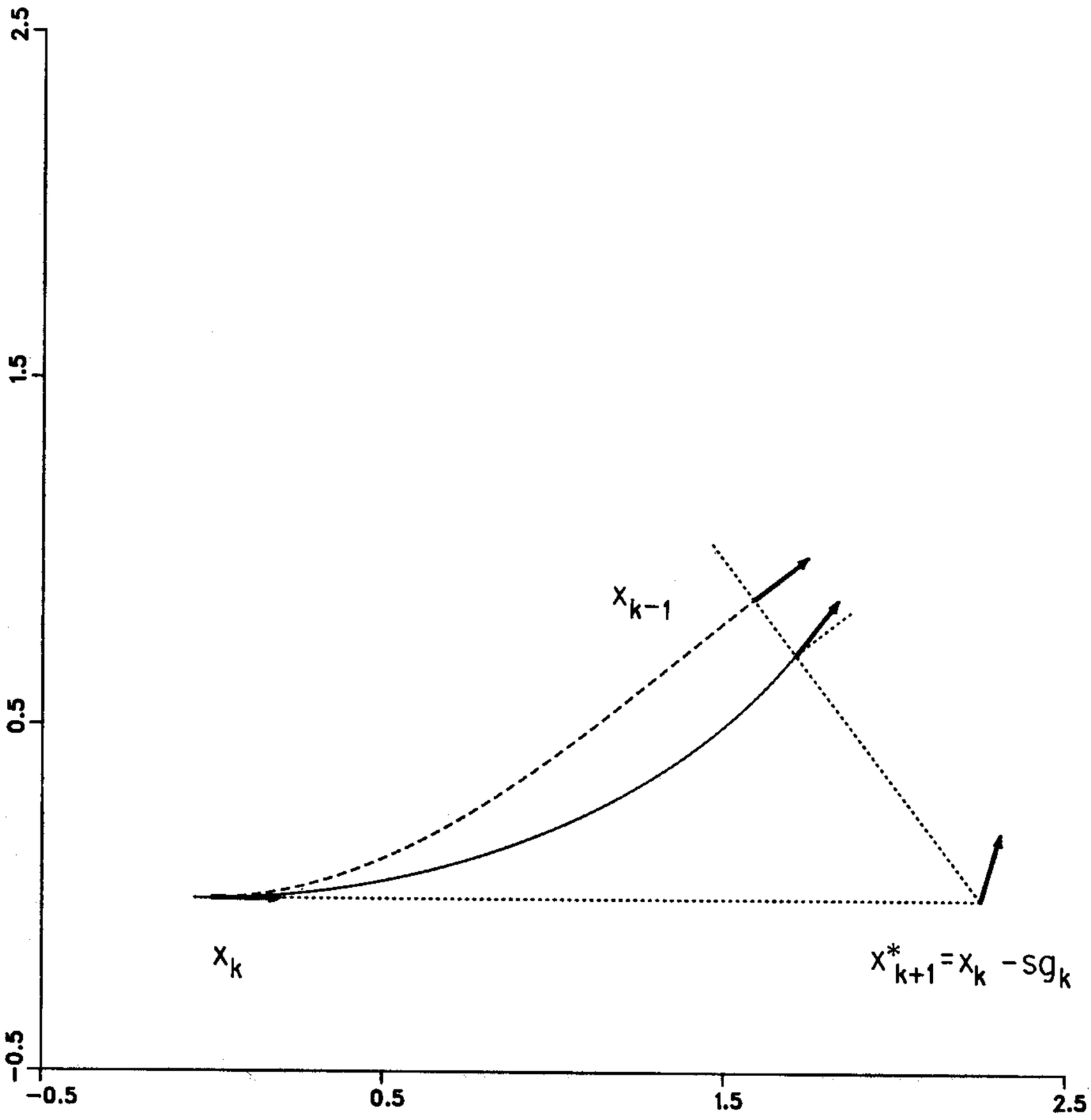

Fig. 10. Reaction path-following algorithm of Ishida, Morokuma and Komornicki (true path (full curve), approximate path (broken curve), vectors point downhill, indicating the orientation of $-\mathbf{g}$). From $\mathbf{x}_k$, a point on the reaction path, a step is taken along $-\mathbf{g}_k$, yielding position $\mathbf{x}^*_{k+1}$ with gradient $\mathbf{g}^*_{k+1}$. Minimization from $\mathbf{x}^*_{k+1}$ along the direction $\mathbf{d}_k = \mathbf{g}^*_{k+1}/|\mathbf{g}^*_{k+1}| - \mathbf{g}_k/|\mathbf{g}_k|$ results in $\mathbf{x}_{k+1}$, the approximation to the next point on the reaction path.

approximation to the solution of the differential equation for the intrinsic reaction path:[13]

$$d\mathbf{x}^{IRC}/ds = -\mathbf{g}_k/|\mathbf{g}_k| \tag{20}$$

A more sophisticated numerical integration of this equation, e.g. a predictor–corrector method, can lead to a much improved reaction path.[118]

An algorithm proposed by Ishida, Morokuma and Komornicki[119] improves on Eq. (19) by adding a one-dimensional search to return the predicted point closer to the IRC (Fig. 10). A step of length s in the direction of $-\mathbf{g}_k$

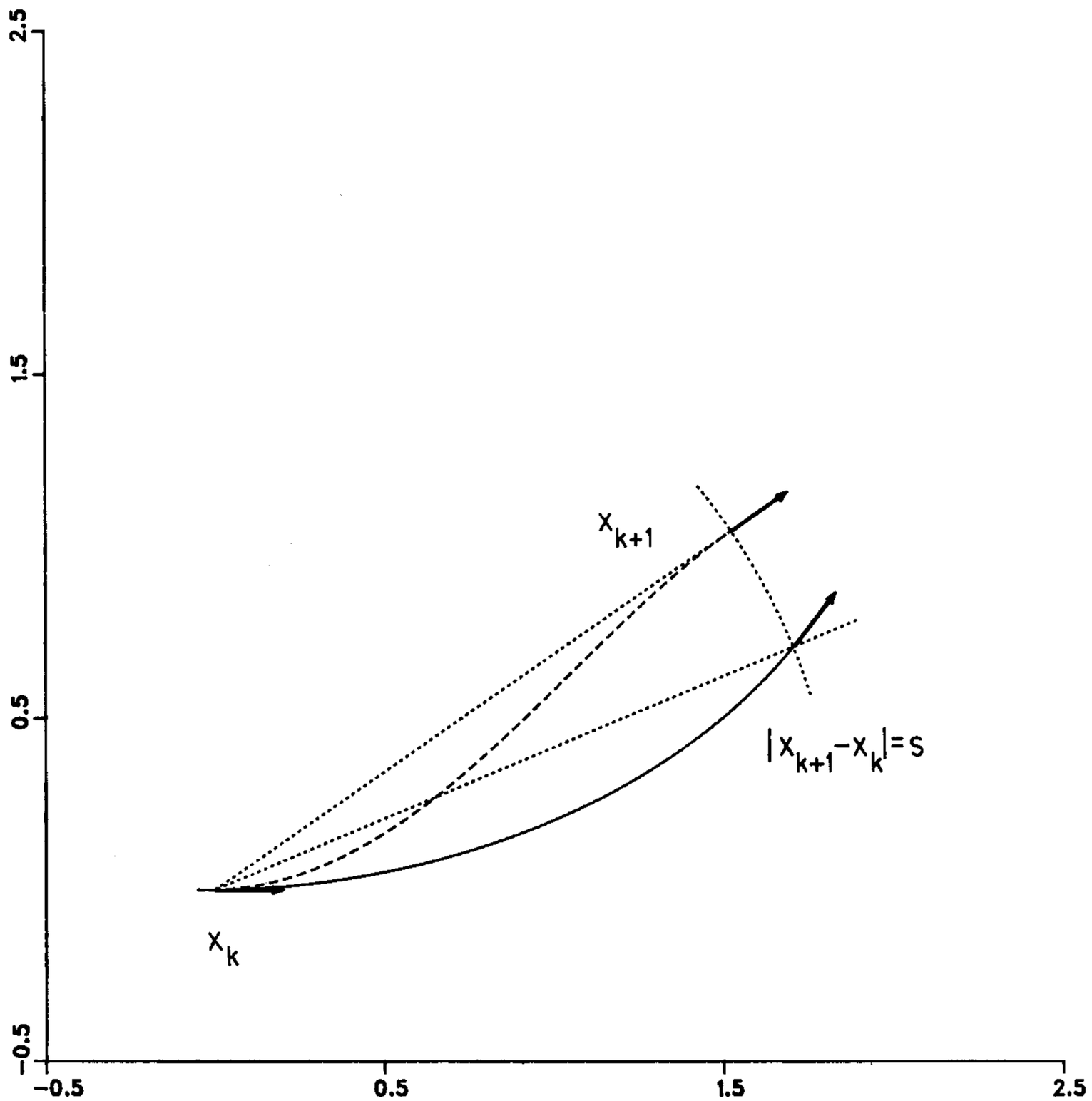

Fig. 11. Reaction path-following algorithm of Müller and Brown (true path (full curve), approximate path (broken curve), vectors point downhill, indicating the orientation of $-\mathbf{g}$). From $\mathbf{x}_k$, a point on the reaction path, a step of length s is taken in the downhill direction. The approximation to the next point on the path, $\mathbf{x}_{k+1}$, is found by minimizing the energy with $\mathbf{x}_{k+1}$ constrained to be on the hypersphere with radius s centered at $\mathbf{x}_k$.

yields point $\mathbf{x}^*_{k+1}$. A minimization along $\mathbf{d}_k$, the bisector of the angle between $-\mathbf{g}_k$ and $\mathbf{g}^*_{k+1}$, results in $\mathbf{x}_{k+1}$, an approximation to the next point on the reaction path. Schmidt, Gordon and Dupuis[120] avoid an explicit search along $\mathbf{d}_k$ by computing only one additional point on $\mathbf{d}_k$ and using a parabolic fit to interpolate to the minimum. A step size of $|\Delta\mathbf{x}/\mathbf{g}| = 0.15$ bohr amu$^{1/2}$ was found to be reasonable and 50–200 energy or energy and gradient calculations were typically needed to map out the IRC for small molecules.[119,120]

In the method of Müller and Brown,[103] a point on the steepest-descent path is found by taking a step of fixed length and minimizing the energy with respect

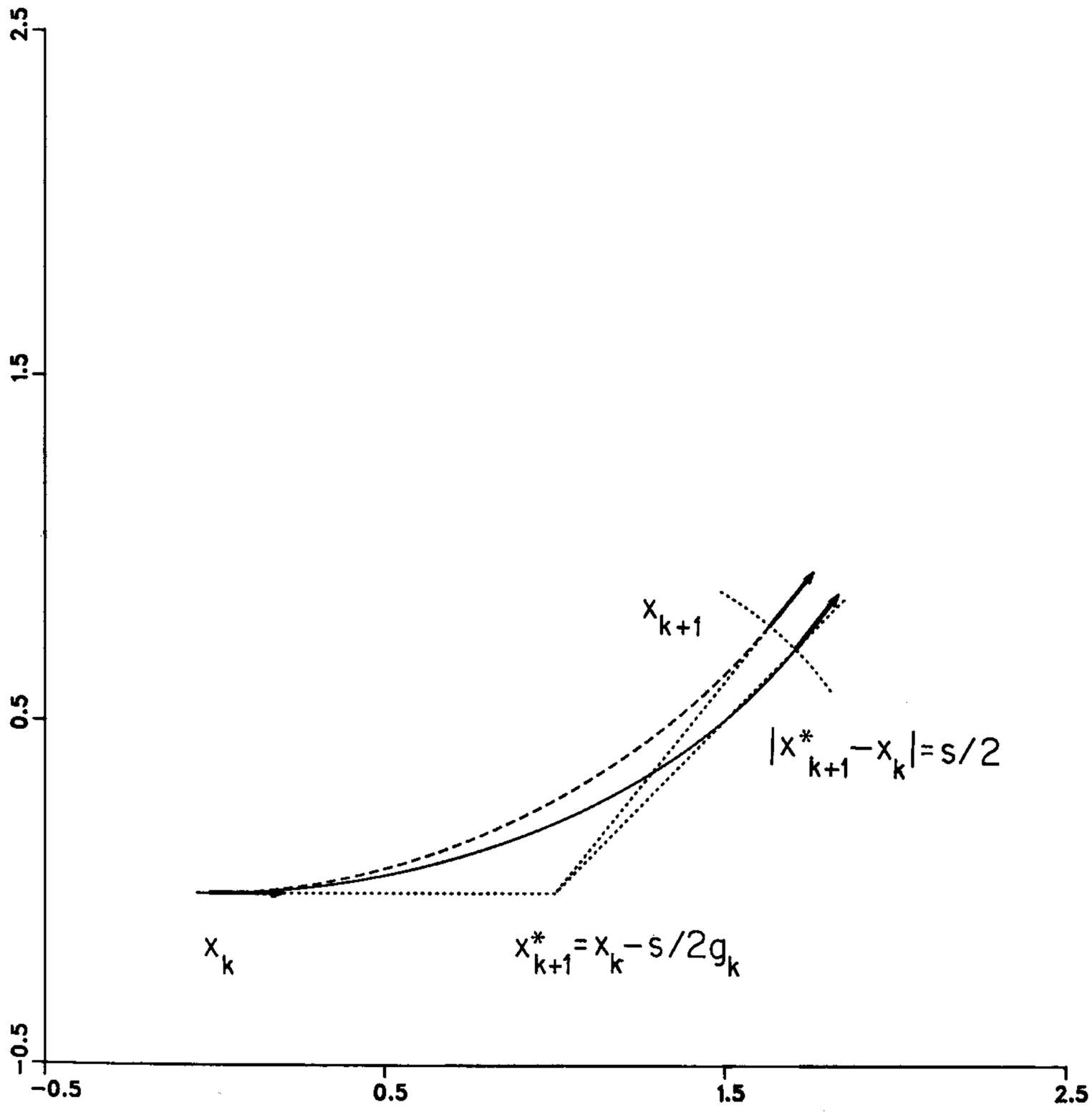

Fig. 12. Reaction path-following algorithm of Schlegel (true path (full curve), approximate path broken curve), vectors point downhill, indicating the orientation of $-\mathbf{g}$). From $\mathbf{x}_k$, a point on the reaction path, a step of length $\frac{1}{2}s$. is taken along $-\mathbf{g}_k$ yielding $\mathbf{x}^*_{k+1}$ (no energy or gradient calculation at $\mathbf{x}^*_{k+1}$). The approximation to the next point on the path, $\mathbf{x}_{k+1}$, is found by minimizing the energy with $\mathbf{x}_{k+1}$ constrained to be on the hypersphere with radius $\frac{1}{2}s$ centered at $\mathbf{x}^*_{k+1}$.

to the remaining $N-1$ degrees of freedom, as shown in Fig. 11. The optimized point $\mathbf{x}_{k+1}$ lies on a hypersphere centered at $\mathbf{x}_k$, and the residual gradient, $\mathbf{g}_{k+1}$, is normal to the sphere (i.e. parallel to $\mathbf{x}_{k+1}-\mathbf{x}_k$). Although each step requires an $(N-1)$-dimensional optimization, a much larger step size can be used than in the methods described above (typically 10 steps from the saddle point to a minimum). For large step sizes and tightly curved reaction paths, this method has a tendency to underestimate the curvature.[103] This is because the residual gradient is parallel to the straight-line path between $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$, rather than tangent to the curved path.

A refinement by Schlegel[121] improves the ability of this approach to follow a curved reaction path, as shown in Fig. 12. The point $\mathbf{x}_{k+1}$ is chosen so that the reaction path between $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ is an arc of a circle and so that the gradients $\mathbf{g}_k$ and $\mathbf{g}_{k+1}$ are tangent to this path. Like the Müller and Brown method, this algorithm requires an $(N-1)$-dimensional optimization on a hypersphere, but about the point $\mathbf{x}^*_{k+1}=\mathbf{x}_k-\frac{1}{2}s\mathbf{g}_k/|\mathbf{g}_k|$ rather than $\mathbf{x}_k$. For similar step sizes, this algorithm follows an IRC with large curvature more closely than the Müller and Brown method.

VII. SUMMARY

This survey of methods as well as the accumulated experience by a number of groups suggest several general statements about geometry optimization. Full geometry optimization should be carried out, if possible, rather than partial optimization, and the stationary points found should be characterized by computing the Hessian. Both of these points are particularly important for transition states. For minima and for saddle points, gradient methods appear to be the most efficient techniques for optimization, especially if analytical gradients are available. Second derivative methods are more suitable for very difficult optimization problems, or if analytical second derivatives are relatively inexpensive. Methods for following intrinsic reactions paths have not been tested as widely and it is not yet possible to recommend one method over another.

Beyond the choice of optimization method, the rate of convergence to the stationary point depends significantly on the choice of coordinate system, the starting coordinates and the initial estimate of the Hessian. Coordinate systems (usually internal coordinates) should avoid strong coupling, especially between stiff and flexible modes, and between variables dominating the transition vector and other coordinates. Good starting geometries can be obtained from the structures of similar molecules found in various bibliographies, or from lower-level calculations. Suitable initial estimates of the Hessian can be calculated by empirical means, by semi-empirical MO methods or by *ab initio* MO methods. For transition structure optimizations, an initial estimate of the transition vector is also needed, either to partition

the space of variables to be optimized, or to insure that the Hessian has a suitable eigenvector with a negative eigenvalue.

References

1. Richards, W. G., Walker, T. E. H., and Hinkley, R. K., *A Bibliography of ab initio Molecular Wavefunctions*, Oxford University Press, Oxford, 1971, and supplements for 1970–73, 1974–77, 1978–80.
2. Whiteside, R. A., Frisch M. J., and Pople, J. A., *The Carnegie–Mellon Quantum Chemistry Archive*, 3rd Ed, Carnegie–Mellon University, Pittsburgh, 1983.
3. Ohno, K., and Morokuma, K., *Quantum Chemistry Literature Data Base*, Elsevier, Amsterdam, 1982; yearly supplements published in special issues of the journal *Theochem.*
4. Bartož, S., *Colloq. Int. CNRS*, **82**, 287 (1958).
5. Bishop, D. M., and Randic, M., *J. Chem. Phys.*, **44**, 2480 (1966).
6. Gerratt, J., and Mills, I. M., *J. Chem. Phys.*, **49**, 1719 (1968).
7. Pulay, P., *Mol. Phys.*, **17**, 197 (1969).
8. Fogarasi, G., and Pulay, P., *Annu. Rev. Phys. Chem.*, **35**, 191 (1984); Pulay, P., in *The Force Concept in Chemistry* (Ed. B. M. Deb), p. 449, Van Nostrand, New York, 1981; Pulay, P., in *Applications of Electronic Structure Theory* (Ed. H. F. Schaefer III), p. 153, Plenum,, New York, 1977.
9. Gaw, J. F., and Handy, N. C., *Annu. Rep. Prog. Chem. Sec. C*, **81**, 291 (1985).
10. Schlegel, H. B., in *Computational Theoretical Organic Chemistry* (Eds. I. G. Csizmadia, and R. Daudel), p. 129, Reidel, Dordrecht, 1981.
11. Jørgensen, P., and Simons, J. (Eds.), *Geometrical Derivatives of Energy Surfaces and Molecular Properties*, Reidel, Dordrecht, 1986.
12. Pulay, P., *Adv. Chem. Phys.*, (1986).
13. Fukui, K., *Acc. Chem. Res.*, **14**, 363 (1981).
14. Tachibana, A., and Fukui, K., *Theor. Chim. Acta*, **51**, 275 (1979).
15. Truhlar, D. G., and Garrett, B. C., *Acc. Chem. Res.*, **13**, 440 (1980).
16. Miller, W. H., Handy, N. C., and Adams, J. E., *J. Chem. Phys.*, **72**, 99 (1980).
17. For leading references see Mezey, P. G., *Theor. Chim. Acta*, **62**, 133 (1982).
18. Longuet-Higgins, H. C., *Mol. Phys.*, **6**, 445 (1963); for a recent review see Frei, H., Bauder, A., and Gunthard, H. H., *Top. Curr. Chem.*, **81**, 1 (1979).
19. Klemperer, W. G., *J. Chem. Phys.*, **56**, 5475 (1972); for recent references see McLarnan, T. J., *Theor. Chim. Acta*, **63**, 195 (1983).
20. Murrell, J. H., and Laidler, K. J., *Trans. Faraday Soc.*, **64**, 371 (1968); Murrell, J. H., and Pratt, G. L., *Trans. Faraday Soc.*, **66**, 1680 (1970).
21. Stanton, R. E., and McIver, Jr, J. W., *J. Am. Chem. Soc.*, **97**, 3632 (1975).
22. Schlegel, H. B., *Theor. Chim. Acta*, **66**, 333 (1984).
23. For a recent review see Hegarty, D., and van der Velde, G., *Int. J. Quantum Chem.*, **23**, 1135 (1983).
24. Saunders, V. R., in *Methods in Computational Molecular Physics*, (Eds. G. H. F., Diercksen, and S. Wilson), p. 1, Reidel, Dordrecht, 1983.
25. Schlegel, H. B., *J. Chem. Phys.*, **77**, 3676 (1982).
26. Schlegel, H. B., Binkley, J. S., and Pople, J. A., *J. Chem. Phys.*, **80**, 1976 (1984).
27. Dupuis, M., and King, H. F., *J. Chem. Phys.*, **68**, 3998 (1978).
28. Takada, T., Dupuis, M., and King, H. F., *J. Chem. Phys.*, **75**, 332 (1981).
29. A partial list of integral derivative codes is included in section of the programs

described is references 30–37 and 60 (these codes have also been adapted for numerous other molecular orbital programs).
30. Binkley, J. S., Whiteside, R. A., Krishnan, R., Seeger, R., DeFrees, D. J., Schlegel, H. B., Topiol, S., Kahn, L. R., and Pople, J. A., GAUSSIAN 80, *Quantum Chem. Program Exch.*, **13**, 406 (1981).
31. Dupuis, M., Rys, J., and King, H. F., HONDO 5, *Quantum Chem. Program Exch.*, **13**, 401, 403 (1981).
32. Komornicki, A., GRADSCF, *NRCC Program* QH04, 1980.
33. Meyer, W., and Pulay, P., MOLPRO, Munich and Stuttgart, Germany, 1969.
34. Pulay, P., TEXAS, *Theor. Chim. Acta*, **50**, 299 (1979).
35. Huber, H., Čársky, P., and Zahradnik, R., POLYGRAD, *Theor. Chim. Acta*, **41**, 217 (1976).
36. Schlegel, H. B., FORCE/DRVEXP, *Quantum Chem. Program Exch.*, **13**, 427 (1981).
37. Amos, R. D., CADPAC, *SERC Daresburg Lab. Publication CCP1/84/4* (1984).
38. Kahn, L. R., *J. Chem. Phys.*, **75**, 3962 (1981).
39. Vincent, M. A., Saxe, P., and Schaefer, III, H. F., *Chem. Phys. Lett.*, **94**, 351 (1983); Vincent, M. A., and Schaefer, III, H. F., *Theor. Chem. Acta*, **64**, 21 (1983).
40. Page, M., Saxe, P., Adams, G. F., and Lengsfield, B. H., *Chem. Phys. Lett.*, **104**, 587 (1984); Banerjee, A., Jensen, J. O., and Simons, J., *J. Chem. Phys.*, **82**, 4566 (1985).
41. Kato, S., and Morokuma, K., *Chem. Phys. Lett.*, **65**, 19 (1979).
42. Goddard, J. D., Handy, N. C., and Schaefer, III, H. F., *J. Chem. Phys.*, **71**, 1525 (1979).
43. Dupuis, M., *J. Chem. Phys.*, **74**, 5758 (1981).
44. Schlegel, H. B., and Robb, M. A., *Chem. Phys. Lett.*, **93**, 43 (1982).
45. Osamura, Y., Yamaguchi, Y., and Schaefer, III, H. F., *J. Chem. Phys.*, **77**, 385 (1982).
46. Pople, J. A., Krishnan, R., Schlegel, H. B., and Binkley, J. S., *Int. J. Quantum Chem. Symp.*, **13**, 225 (1979).
47. Saxe, P., Yamaguchi, Y., and Schaefer, III, H. F., *J. Chem. Phys.*, **77**, 5647 (1982).
48. Osamura, Y., Yamaguchi, Y., Saxe, P., Vincent, M. A., Gaw, J. F., and Schaefer, III, H. F., *Chem. Phys.*, **72**, 131 (1982).
49. Pulay, P., *J. Chem. Phys.*, **78**, 5043 (1983).
50. Jørgensen, P., and Simons, J., *J. Chem. Phys.*, **79**, 334 (1983).
51. Almöf, J., and Taylor, P. R., *Int. J. Quantum Chem.*, **27**, 743 (1985).
52. Yamaguchi, Y., Osamura, Y., Fitzgerald, G., and Schaefer, III, H. F., *J. Chem. Phys.*, **78**, 1607 (1983).
53. Camp, R. N., King, H. F., McIver, Jr, J. W., and Mullally, D., *J. Chem. Phys.*, **79**, 1089 (1983).
54. Hoffmann, M. R., Fox, D. J., Gaw, J. F., Osamura, Y., Yamaguchi, Y., Grev, R. S., Fitzgerald, G., Schaefer, III, H. F., Knowles, P. J., and Handy, N. C., *J. Chem. Phys.*, **80**, 2660 (1984).
55. Page, M., Saxe, P., Adams, G. F., and Lengsfield, III, B. H., *J. Chem. Phys.*, **81**, 434 (1984).
56. Banerjee, A., Jensen, J. O., Simons, J., and Shepard, R., *Chem. Phys.*, **87**, 203 (1984).
57. Binkley, J. S., unpublished.
58. Osamura, Y., Yamaguchi, Y., Saxe, P., Fox, D. J., Vincent, M. A., and Schaefer, III, H. F., *J. Mol. Struct.*, **103**, 183 (1983).
59. Simons, J., and Jørgensen, P., *J. Chem. Phys.*, **79**, 3599 (1983).
60. Gaw, J. F., Yamaguchi, Y., and Schaefer, III, H. F., *J. Chem. Phys.*, **81**, 6395 (1984).
61. Handy, N. C., and Schaefer, III, H. F., *J. Chem. Phys.*, **81**, 5031 (1984).

62. Krishnan, R., Schlegel, H. B., and Pople, J. A., *J. Chem. Phys.*, **72**, 4654 (1980).
63. Brooks, B. R., Laidig, W. D., Saxe, P., Goddard, J. D., Yamaguchi, Y., and Schaefer, III, H. F., *J. Chem. Phys.*, **72**, 4652 (1980).
64. Osamura, Y., Yamaguchi, Y., and Schaefer, III, H. F., *J. Chem. Phys.*, **75**, 2919 (1981).
65. Fitzgerald, G., Harrison, R., Ladig, W. D., and Bartlett, R. J., *J. Chem. Phys.*, **82**, 4379 (1985).
66. Fitzgerald, G., Harrison, R., Ladig, W. D., and Bartlett, R. J., *Chem. Phys. Lett.*, **117**, 433 (1985).
67. Pulay, P., *J. Mol. Struct.*, **103**, 57 (1983).
68. Fox, D. J., Osamura, Y., Hoffmann, M. R., Gaw, J. F., Fitzgerald, G., Yamaguchi, Y. and Schaefer, III, H. F., *Chem. Phys. Lett.*, **102**, 17 (1983).
69. Bartlett, R. J., Harrison, R. A., Fitzgerald, G., and Ladig, W. D., to be published.
70. For discussions for non-linear optimization methods see references 71–74 and related books.
71. Fletcher, R., *Practical Methods, of Optimization*, Wiley, Chichester, 1981.
72. Gill, P. E., Murray, W., and Wright, M. H., *Practical Optimization*, Academic Press, New York, 1981.
73. Powell, M. J. D. (Ed.), *Non-linear Optimization, 1981*, Academic Press, New York, 1982.
74. Scales, L. E., *Introduction to Non-linear Optimization*, Macmillan, Basingstoke, 1985.
75. Fletcher, R., and Powell, M. J. D., *Comput. J.*, **6**, 163 (63). Davidon, W., *Argonne National Lab. Report*, ANL-5990.
76. Binkley, J. S., *J. Chem. Phys.*, **64**, 5142 (1976).
77. Payne, P. W., *J. Chem. Phys.*, **65**, 1920 (1976).
78. Sana, M., *Int. J. Quantum Chem.*, **19**, 139 (1981); Comeau, D. C., Zellmer, R. J., and Shavitt, I., in Ref. 11.
79. Fletcher, R., and Reeves, C. M., *Comput. J.*, **7**, 149 (1964).
80. Murtagh, B. A., and Sargent, R. W. H., *Comput. J.*, **13**, 185 (1972).
81. Broyden, C. G., *J. Inst. Math. Appl.*, **6**, 76 (1970); Fletcher, R., *Comput. J.*, **13**, 317 (1970); Goldfarb, D., *Math. Comput.*, **24**, 23 (1970); Shanno, D. F., *Math. Comput.*, **24**, 647 (1970).
82. Davidon, W. C., *Math. Programming*, **9**, 1 (1975).
83. Broyden, C. G., *Math. Comput.*, **19**, 368 (1967).
84. Huang, D. Y., *J. Opt. Theor. Appl.*, **5**, 405 (1970).
85. Biggs, M. C., *J. Inst. Math. Appl.*, **12**, 337 (1973).
86. Schlegel, H. B., *J. Comput. Chem.*, **3**, 214 (1982).
87. Davidon, W. C., in Ref. 73, p. 23.
88. McKelvey, J. M., and Hamilton, Jr, J. F., *J. Chem. Phys.*, **80**, 579 (1984).
89. Császár, P., and Pulay, P., *J. Mol. Struct.*, **114**, 31 (1984).
90. Badger, R. M., *J. Chem. Phys.*, **2**, 128 (1934); *ibid.*, **3**, 227 (1935).
91. For other reviews of transition structure optimization algorithms see Refs. 92 and 93.
92. Müller, K., *Angew. Chem. Int. Edn, Engl.*, **19**, 1 (1980).
93. Bell, S., and Crighton, J. S., *J. Chem. Phys.*, **80**, 2464 (1984).
94. For leading references see Truhlar, D. G., (Ed.), *Potential Energy Surfaces and Dynamics Calculations*, Plenum, New York, 1981.
95. Halgren, T. A., and Lipscomb, W. N., *Chem. Phys. Lett.*, **49**, 225 (1977).
96. Scharfenberger, P. *J. Comput. Chem.*, **3**, 277 (1982).
97. Bálint, I., and Bán, M. I., *Theor. Chim. Acta*, **63**, 255 (1983).

98. Jensen, A., *Theor. Chim. Acta*, **63**, 269 (1983).
99. Tapia, O., and Andrés, J., *Chem. Phys. Lett.*, **109**, 471 (1984).
100. Bell, S., Crighton, J. S., and Fletcher, R., *Chem. Phys. Lett.*, **82**, 122 (1981).
101. Rothman, M. J., and Lohr, Jr, L. L., *Chem. Phys. Lett.*, **70**, 405 (1980).
102. Burkert, U., and Allinger, N. L., *J. Comput. Chem.*, **3**, 40 (1982).
103. Müller, K., and Brown, L. D., *Theor. Chim. Acta*, **53**, 75 (1979).
104. Basilevsky, M. V., and Shamov, A. G., *Chem. Phys.*, **60**, 347 (1981).
105. Dewar, M. J. S., Healy, E. F., and Stewart, J. J. P., *J. Chem. Soc. Faraday Trans. 2*, **80**, 227 (1984).
106. Cerjan, C. J., and Miller, W. H., *J. Chem. Phys.*, **75**, 2800 (1981).
107. Simons, J., Jørgensen, P., Taylor, H., and Ozment, J., *J. Phys. Chem.*, **87**, 2745 (1983).
108. Banerjee, A., Adams, N., Simons, J., and Shepard, R., *J. Phys. Chem.*, **89**, 52 (1985).
109. McIver, Jr, J. W., and Komornicki, A., *J. Am. Chem. Soc.*, **94**, 2625 (1972).
110. Poppinger, D., *Chem. Phys. Lett.*, **35**, 550 (1975).
111. Komornicki, A., Ishida, K., Morokuma, K., Ditchfield, R., and Conrad, M., *Chem. Phys. Lett.*, **45**, 595 (1977).
112. Powell, M. J. D., *Comput. J.*, **7**, 303 (1965).
113. For leading references see Franke, R., *Math. Comput.*, **38**, 181 (1982).
114. Shepard, D., *Proc ACM Natl Conf.*, p. 517 (1968).
115. Schlegel, H. B., to be published.
116. Sana, M., Reckinger, G., and Leroy, G., *Theor. Chim. Acta*, **58**, 145 (1981); Quapp, W., and Heidrich, D., *Theor. Chim. Acta*, **66**, 245 (1984).
117. Baskin, C. P., Bender, C. F., Bauschlicher, Jr, C. W., and Schaefer, III, H. F., *J. Am. Chem. Soc.*, **96**, 2709 (1974).
118. Hase, W. L., unpublished.
119. Ishida, K., Morokuma, K., and Komornicki, A., *J. Chem. Phys.*, **66**, 2153 (1977).
120. Schmidt, M. W., Gordon, M. S., and Dupuis, M., *J. Am. Chem. Soc.*, **107**, 2585 (1985).
121. Schlegel, H. B., to be published.