

95. Y. Yokobayashi, K. Ikebukuro, S. McNiven, and I. Karube, *J. Chem. Soc., Perkin Trans. J.*, 1996, 2435-2437.
96. J. Mestres and G. E. Scuseria, *J. Comput. Chem.*, 1995, **16**, 729-742.
97. S. K. Gregurick, M. H. Alexander, and B. Hartke, *J. Chem. Phys.*, 1996, **104**, 2684-2691.
98. J. A. Niesse and H. R. Mayne, *J. Chem. Phys.*, 1996, **105**, 4700-4706.
99. D. M. Deaven and K. M. Ho, *Phys. Rev. Lett.*, 1995, **75**, 288-291.
100. D. M. Deaven, N. Tit, J. R. Morris, and K. M. Ho, *Chem. Phys. Lett.*, 1996, **256**, 195-200.
101. A. Tomasulo and M. V. Ramakrishna, *J. Phys. Chem.*, 1996, **105**, 10449-10455.
102. D. Rogers and A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 854-866.
103. R. Leardi, *J. Chemometr.*, 1994, **8**, 65-79.
104. H. Kubinyi, *Quant. Struct.-Act. Relat.*, 1994, **13**, 285-294.
105. H. Kubinyi, *Quant. Struct.-Act. Relat.*, 1994, **13**, 393-401.
106. H. Kubinyi, *J. Chemometr.*, 1996, **10**, 119-133.
107. S. S. So and M. Karplus, *J. Med. Chem.*, 1996, **39**, 1521-1530.
108. J. Kyngäs and J. Valjakka, *Quant. Struct.-Act. Relat.*, 1996, **15**, 296-301.
109. G. Chang and M. Lewis, *Acta Crystallogr.*, 1994, **D50**, 667-674.
110. K. Y. Tam and R. G. Compton, *J. Appl. Crystallogr.*, 1995, **28**, 640-645.
111. T. S. Bush, C. R. A. Catlow, and P. D. Battle, *J. Mater. Chem.*, 1995, **5**, 1269-1272.
112. S. T. Miller, J. M. Hogle, and D. J. Filman, *Acta Crystallogr.*, 1996, **D52**, 235-251.
113. A. W. R. Payne and R. C. Glen, *J. Mol. Graphics*, 1993, **11**, 74-91.
114. G. Jones, P. Willett, and R. C. Glen, *J. Comput.-Aided Mol. Des.*, 1995, **9**, 532-549.
115. D. E. Walters and R. M. Hinds, *J. Med. Chem.*, 1994, **37**, 2527-2536.
116. Y. Zeiri, E. Fattal, and R. Kosloff, *J. Chem. Phys.*, 1992, **96**, 1859-1862.
117. I. Rossi and D. G. Truhlar, *Chem. Phys. Lett.*, 1995, **233**, 231-236.
118. J. H. Rodriguez, H. N. Ok, Y.-M. Xia, P. G. Debrunner, B. E. Hinrichs, T. Meyer, and N. H. Packard, *J. Phys. Chem.*, 1996, **100**, 6849-6862.
119. A. C. W. May and M. S. Johnson, *Protein Eng.*, 1994, **7**, 475-485.
120. A. C. W. May and M. S. Johnson, *Protein Eng.*, 1995, **8**, 873-882.
121. E. P. Jaeger, D. C. Pevear, P. J. Felock, G. R. Russo, and A. M. Treasurywala, *ACS Symp. Ser.*, 1995, **589**, 139-155.
122. R. E. Shaffer, G. W. Small, and M. A. Arnold, *Anal. Chem.*, 1996, **68**, 2663-2675.
123. A. S. Bangalore, R. E. Shaffer, G. W. Small, and M. A. Arnold, *Anal. Chem.*, 1996, **68**, 4200-4212.
124. P. Vankeerberghen, J. Smeyers-Verbeke, R. Leardi, C. L. Karr, and D. L. Massart, *Chemometr. Intell. Lab. Syst.*, 1995, **28**, 73.
125. B. M. Wise, B. R. Holt, N. B. Gallager, and S. Lee, *Chemometr. Intell. Lab. Syst.*, 1995, **30**, 81-89.
126. A. H. C. van Kampen, C. S. Strom, and L. M. C. Buydens, *Chemometr. Intell. Lab. Syst.*, 1996, **34**, 55-68.
127. A. Gilman and G. Ross, *Biophys. J.*, 1995, **69**, 1321-1333.
128. J. Dods, D. Gruner, and P. Brumer, *Chem. Phys. Lett.*, 1996, **261**, 612-619.
129. A. D. Dane, P. A. M. Timmermans, H. A. van Sprang, and L. M. C. Buydens, *Anal. Chem.*, 1996, **68**, 2419-2425.
130. D. B. Hibbert, *Chemometr. Intell. Lab. Syst.*, 1993, **19**, 319-329.
131. R. Moros, H. Kalies, H. G. Rex, and S. Schaffarczyk, *Comput. Chem. Eng.*, 1996, **20**, 1257-1270.

Geometry Optimization: 1

H. Bernhard Schlegel

Wayne State University, Detroit, MI, USA

1	Introduction	1136
2	Methods for Finding Local Minima	1138
3	Methods for Finding Transition Structures	1140
4	Related Articles	1142
5	References	1142

Abbreviations

BFGS = Broyden-Fletcher-Goldfarb-Shanno; DFP = Davidson-Fletcher-Powell; EF = eigenvector following; GDIIS = geometry optimization by direct inversion of the iterative subspace; LST = linear synchronous transit; QST = quadratic synchronous transit; RFO = rational function optimization.

1 INTRODUCTION

The concept of a potential energy surface is central to the discussion of molecular structures and geometry optimization. The potential energy surface describes the energy of the molecule as a function of its geometry (i.e., bond lengths, valence angles, torsions, and other internal coordinates). Equilibrium geometries are local minima on the potential energy surface for a molecule; transition states correspond to saddle points on the potential energy surface. Geometry optimization is the process of finding these minima and saddle points. Because molecules may contain many atoms and have many coordinates to describe the molecular geometry, it is difficult to calculate or depict the entire potential energy surface. However, a simplified two-dimensional potential energy surface can be visualized as a hilly landscape with valleys representing reactants, intermediates, and products; transition states correspond to mountain passes connecting the valleys. Figure 1 shows a simple two-dimensional potential energy surface, and illustrates some of its features.

Potential energy surfaces arise naturally from the Born-Oppenheimer approximation. Because the electrons are so much lighter than the nuclei, the electronic part of the wavefunction can readjust almost instantaneously to any nuclear motion. In the Born-Oppenheimer approximation, a potential energy surface is obtained by solving for the electronic energy at a series of fixed nuclear positions. This is usually quite satisfactory for most ground state systems. However, for photochemical systems, which involve excited state surfaces as well as the ground state, one must go beyond the Born-Oppenheimer approximation to treat molecular motion near seams, conical intersections, and weakly avoided crossings (see *Nonadiabatic Derivative Couplings and Photochemistry*).

Potential energy surfaces make it possible to discuss molecular structures. The equilibrium geometry of a molecule corresponds to a minimum on the potential energy surface.

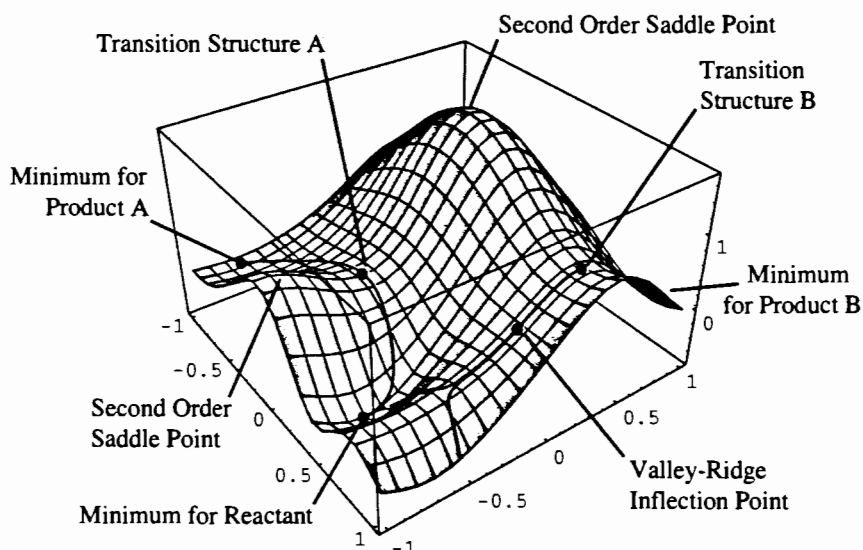


Figure 1 A model potential energy surface illustrating minima, transition structures, second-order saddle points, a valley-ridge inflection point, and reaction paths (Reproduced from Ref. 2a, with permission from World Scientific Publishing)

There may be several minima, representing different conformers and isomers of the molecule, or representing reactants, intermediates, and products of a chemical reaction. If the valley on the potential energy surface is deep with steep sides, then the structure of the molecule is fairly rigid and well defined. However, if the valley is broad or shallow, the molecule is flexible or reactive, and the concept of molecular structure is less well defined.

A reaction can be represented by a movement from the valley of the reactants to the valley of the products. Depending on the nature of the reactants and products, the reaction may be a simple conformational change, an isomerization, a unimolecular reaction, or a bimolecular reaction. The particular path across the potential energy surface describes the reaction mechanism. The transition structure is the highest point on that reaction path which requires the least energy to get from the reactants to the products. Its energy, structure and vibrational frequencies can be used to predict reaction rates by transition state theory (see *Transition State Theory*). Both minima and transition structure can be located by geometry optimization. Efficient methods are also available for following reaction paths once the transition structure has been located (see *Reaction Path Following*).

Since accurate functions for potential energy surfaces are difficult to obtain even for very small systems, most geometry optimization methods find equilibrium structures and transition structures directly, without constructing the full potential energy surface. Efficient geometry optimization methods employ the first derivatives of the potential energy surfaces with respect to the geometric parameters. Usually these are obtained analytically (see *Gradient Theory*), but for theoretical methods without analytical derivatives they can be calculated numerically. For some algorithms, second derivatives can also be used.

The first derivatives of the potential energy surface are also called the gradient. In classical mechanics, the negatives of the first derivatives of the potential are the forces on the atoms in the molecule, $-\partial V/\partial x = F_x$. Thus, points on the potential energy surface where the gradient or forces are

zero are called stationary points. In a topological analysis of the potential energy surface, these points are known as critical points. The matrix of second derivatives of the potential energy surface is termed the *Hessian matrix* or the harmonic force constant matrix. At a critical point, diagonalization of the mass-weighted force constant matrix yields the vibrational frequencies and normal modes. The number of negative eigenvalues of the Hessian or the number of imaginary frequencies at a stationary or critical point is known as the index of the critical point. A critical point of index 0 is a minimum and a critical point of index 1 is a transition structure.

For a point on the potential energy surface to be a minimum, it must satisfy two conditions. The first derivatives, or equivalently the gradient or the forces, must be zero (i.e., they must be a stationary or critical point). If the first derivatives are not zero, there is a nearby point that is lower in energy. Secondly, the second derivative matrix or Hessian or force constant matrix must be positive definite (i.e., index 0). In other words, all the eigenvalues of the Hessian must be positive for a minimum or all of the vibrational frequencies must be real (i.e., no imaginary frequencies). Note that overall translation and rotation of the molecule and any redundancy in the internal coordinates must be removed before obtaining the eigenvalues of the Hessian. If one or more eigenvalues are negative, then the potential surface is a maximum along these directions and the point is not a minimum but a saddle point. Figure 1 shows a number of minima.

For a point to be a transition structure, the first derivatives must be zero and the energy must be a maximum along the reaction path connecting reactants and products. The point must also be a minimum for all directions perpendicular to the path. If it is not a minimum in these directions, there is a nearby path with a lower barrier and a lower-energy transition structure. Hence, a transition structure is a critical point of index 1, i.e., one of the eigenvalues of the Hessian must be negative and all the other eigenvalues must be positive. Equivalently, a transition structure must have one and only one imaginary frequency; all the other frequencies must be real. The eigenvector of the Hessian corresponding

to the negative eigenvalue, or the normal mode of vibration corresponding to the imaginary frequency, represents motion along the reaction path and is termed the transition vector. A transition structure is also called a first-order *saddle point*. An n -th-order saddle point has n negative eigenvalues of the Hessian, or n imaginary frequencies. This corresponds to a point that is a maximum in n directions, and hence does not represent a transition structure. Figure 1 shows a number of transition structures and second-order saddle points.

2 METHODS FOR FINDING LOCAL MINIMA

Finding an equilibrium geometry involves an unconstrained optimization on the potential energy surface. Algorithms for minimization of nonlinear functions of many variables are discussed extensively in the numerical analysis literature.¹ For the present discussion these can be grouped into three broad categories: methods that use only the energy (such as univariate search and simplex), gradient methods (conjugate gradient, quasi-Newton, etc.), and algorithms that require second derivatives (Newton or Newton-Raphson methods). Energy-only methods are most widely applicable, but are the slowest to converge. Second derivative methods converge very fast, but analytic second derivatives are available only for a limited number of theoretical methods and are significantly more costly than gradient calculations. Gradient-based algorithms are the methods of choice for most levels of theory. If analytical gradients are not available, it is usually more efficient to calculate them numerically, rather than to use an energy-only optimization method. Algorithms for geometry optimization have been reviewed in a number of recent articles² (see also *Geometry Optimization: 2*).

In Newton and quasi-Newton methods, the potential energy surface is approximated as a quadratic function:

$$E(\bar{x}) = E_0 + \bar{g}_0^T \Delta \bar{x} + (1/2) \Delta \bar{x}^T \mathbf{H} \Delta \bar{x}$$

$$\bar{g}(\bar{x}) = \bar{g}_0 + \mathbf{H} \Delta \bar{x} \quad (1)$$

where $\Delta \bar{x} = \bar{x} - \bar{x}_0$, \bar{g} is the gradient or first derivative, and \mathbf{H} is the Hessian or second derivative. At the minimum, the gradient is zero and the step to the minimum can be obtained by solving a set of linear equations:

$$\bar{g}(\bar{x}) = \bar{g}_0 + \mathbf{H} \Delta \bar{x} = 0$$

$$\Delta \bar{x} = -\mathbf{H}^{-1} \bar{g}_0 \quad (2)$$

This is the Newton step and is the basis of most gradient optimization methods. A flowchart of a typical quasi-Newton optimization method is shown in Figure 2.

The efficiency of a quasi-Newton type geometry optimization depends on six factors: (a) the initial geometry, (b) the coordinate system, (c) the initial guess for the Hessian, (d) the line search, (e) the Hessian updating method, and (f) step size control. These factors are discussed in the following paragraphs.

Obviously, the closer the initial guess to the final optimized geometry, the fewer the steps that will be needed. Good initial structures for electronic structure calculations can be obtained using graphical user interfaces and molecular modeling programs, from molecular mechanics calculations, or from lower-level electronic structure calculations. For molecules with

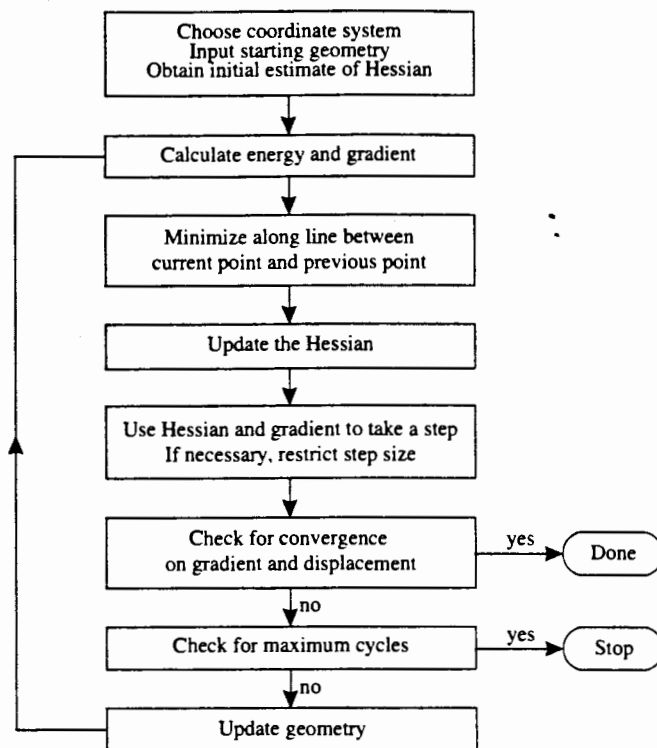


Figure 2 Flowchart for quasi-Newton algorithms for geometry optimization (Reproduced from Ref. 2a, with permission from World Scientific Publishing)

multiple minima, the initial structure should be close to the desired local minimum. Finding the global minimum is a much more difficult problem (see *Genetic Algorithms: Introduction and Applications; Genetic and Evolutionary Algorithms; Macromolecular Structure Calculation and Refinement by Simulated Annealing: Methods and Applications; and Simulated Annealing*) that is outside the scope of this discussion.³

The coordinate system can have a very large effect on the rate of convergence of a geometry optimization. Cartesian coordinates are the simplest and provide an unambiguous representation for any structure. However, Cartesian coordinates are strongly coupled; for example, to change the length of a bond, one has to change the x , y , and z coordinates of two or more atoms. An alternative is to use internal coordinates, such as bond lengths, valence angles, dihedral angles, etc. For acyclic molecules, it is easy to construct a nonredundant internal coordinate system, i.e., one that contains only $3N - 6$ internal coordinates for an N -atom molecule. For a cyclic molecule, a coordinate system that contains all of the bonds and angles would have more than the requisite $3N - 6$ coordinates. For example, a six-membered ring has 6 bonds, 6 valence angles, and 6 dihedrals for a total of 18 coordinates, whereas there are only 12 nonredundant internal degrees of freedom. For molecular systems, optimizations in redundant or natural internal coordinates appear to require significantly fewer optimization steps than nonredundant internal coordinates or Cartesian coordinates.⁴ The transformation of Cartesian energy derivatives to redundant internal coordinates and of redundant internal displacements to Cartesians can be carried out using a generalized inverse.⁴

The initial estimate of the Hessian will also affect the rate of convergence. The closer the guess is to the correct Hessian,

the faster the convergence. Cartesian coordinates for molecules have strong anharmonic coupling between coordinates and the Hessian has many off-diagonal elements. However, if a good estimate of the Hessian is available, the rate of convergence of optimizations in Cartesian coordinates can be quite satisfactory. In internal coordinates, the coupling is much smaller and the Hessian is more nearly diagonal. For cyclic molecules, simple, nonredundant internal coordinates (e.g., coordinates that do not include ring closure bonds) are also strongly coupled, but for redundant or natural internal coordinates the coupling is much smaller and the Hessian can be nearly diagonal. An initial empirical estimate of a diagonal Hessian can be quite satisfactory for redundant internal coordinates and can be readily transformed to other coordinate systems. The initial Hessian for electronic structure calculations can also be computed by molecular mechanics or semiempirical methods. The Hessian from an optimization or a frequency calculation at a lower level of theory is often a very good initial guess for an optimization at a higher level. Some difficult optimizations may require an accurate initial Hessian computed analytically or numerically at the same level of theory as that used for the optimization.

As illustrated in Figure 2, the main loop of a quasi-Newton geometry optimization consists of (i) an energy and gradient evaluation, (ii) a line search, (iii) the Hessian update, (iv) the prediction of the change in geometry, and (v) a convergence test. A one-dimensional minimization along the line between the current point and the previous point can improve the optimization by taking into account the nonquadratic nature of the potential energy surface. An accurate line search would require additional calculations, but an approximate search can be carried out without any extra energy or gradient calculations. A cubic or quartic polynomial can be fitted to the energy and directed gradient at the current and the previous point, and the minimum for the line search can be found using the polynomial.

The next step in a quasi-Newton optimization is the update of the Hessian, from \mathbf{H}^{old} to \mathbf{H}^{new} . We wish to adjust the Hessian in the quadratic approximation represented by equation (1) so that it fits the gradient \bar{g}_i at the current point \bar{x}_i and the gradient \bar{g}_{i-1} at the previous point. This requirement leads to:

$$\mathbf{H}^{\text{new}} \Delta \bar{x} = \Delta \bar{g} \quad (3)$$

where $\Delta \bar{x} = \bar{x}_i - \bar{x}_{i-1}$ and $\Delta \bar{g} = \bar{g}_i - \bar{g}_{i-1}$. There are numerous methods for updating \mathbf{H}^{old} to give \mathbf{H}^{new} ; some of these methods are Murtagh-Sargent, symmetric Powell, Davidson-Fletcher-Powell (DFP), optimally conditioned, and Broyden-Fletcher-Goldfarb-Shanno (BFGS).¹ This latter update is generally accepted as the best formula for minimization, and can be written as:

$$\mathbf{H}^{\text{new}} = \mathbf{H}^{\text{old}} + \Delta \bar{g} \Delta \bar{g}^t / \Delta \bar{x}^t \Delta \bar{g} - \mathbf{H}^{\text{old}} \Delta \bar{x} \Delta \bar{x}^t \mathbf{H}^{\text{old}} / \Delta \bar{x}^t \mathbf{H}^{\text{old}} \Delta \bar{x} \quad (4)$$

This update is symmetric and positive definite, and minimizes the norm of the change in the Hessian. Corresponding formulas to update the inverse of the Hessian are also available.¹

The quasi-Newton geometry optimization methods and related approaches discussed above are suitable for finding minima for small to medium-sized molecules. For difficult cases, the Hessian can be recalculated every few steps or at each step, rather than by an updating scheme. This is equivalent to a Newton or Newton-Raphson algorithm.

Once the Hessian has been updated, a Newton step given by equation (2) is taken on the model quadratic surface. For this step to be in the downhill direction, the Hessian must be positive definite, i.e., all of the eigenvalues must be positive. Positive definite character can be tested and enforced by using a modified Cholesky decomposition to calculate the inverse, or by diagonalizing the Hessian and adjusting the offending eigenvalues.

If the structure is far from the minimum (e.g., large gradients) or the potential energy surface is very flat (one or more small eigenvalues of the Hessian), then a simple Newton step may be too large, taking the molecule beyond the region where the model quadratic surface is valid. Under these circumstances, it is wise to take a shorter step.^{1,2} One can limit the step to be no larger than a trust radius τ . Minimizing the energy in equation (1) subject to the constraint that $|\Delta \bar{x}| \leq \tau$ yields a step equal to:

$$\Delta \bar{x} = -(\mathbf{H} - \lambda \mathbf{I})^{-1} \bar{g}_0 \quad (5)$$

where $\lambda \leq 0$ and λ is less than the lowest eigenvalue of the Hessian; λ is adjusted so that the constraint is satisfied. If b_i are the eigenvalues of the Hessian and \bar{v}_i are the corresponding eigenvectors, the trust radius limited step can also be written as:

$$\begin{aligned} \bar{x}_{i+1} = \bar{x}_i - \sum_{j \geq 1} \frac{\bar{v}_j (\bar{v}_j^t \bar{g}_i)}{(b_j - \lambda)}; \lambda < \min(b_1, 0) \text{ and} \\ \sum_{j \geq 1} \frac{(\bar{v}_j^t \bar{g}_i)^2}{(b_j - \lambda)^2} = \tau^2 \end{aligned} \quad (6)$$

As the optimization proceeds, the trust radius can be increased or decreased, depending on whether the change in the energy predicted by the model surface compares well or poorly with the actual calculated energy difference.

The rational function optimization (RFO) method⁵ of controlling the step size involves minimizing a rational polynomial approximation of the surface:

$$E(\bar{x}) = E_0 + [\bar{g}_0^t \Delta \bar{x} + (1/2) \Delta \bar{x}^t \mathbf{H} \Delta \bar{x}] / (1 + \alpha \Delta \bar{x}^t \Delta \bar{x}) \quad (7)$$

rather than a model quadratic surface, equation (1). The parameter α is adjusted so that the step is downhill and less than or equal to the trust radius τ . This leads to a set of equations equivalent to equations (5) and (6). An advantage of both the trust radius approach and the RFO method is that they will step in the downhill direction even when the Hessian is not positive definite, whereas the simple Newton method (equation 2), will step toward a saddle point if the Hessian is not positive definite.

The GDIIIS method⁶ (geometry optimization by direct inversion of the iterative subspace) is an alternative approach for predicting the change in the geometry that is comparable in efficiency to the quasi-Newton methods. A linear combination of the current and previous points is chosen so that the Newton step is a minimum:

$$\begin{aligned} \bar{x}'_i = \sum c_j \bar{x}_j \text{ and } \bar{x}_{i+1} - \bar{x}'_i = \mathbf{H}^{-1} \sum c_j \bar{g}_j \text{ such that} \\ \sum c_j = 1 \text{ and } |\bar{x}_{i+1} - \bar{x}'_i| \text{ is a minimum} \end{aligned} \quad (8)$$

The GDIIIS approach does not depend as critically on the quality of the Hessian as quasi-Newton methods. It can be

used with a constant Hessian or with updating. If the Hessian is diagonal or a constant and a sparse inverse Hessian is used, the GDIIS method can also be quite efficient for large systems.

For very large molecules, the storage and diagonalization or inversion of the full Hessian can become too costly. One alternative is to use a sparse or blocked Hessian in a quasi-Newton scheme.^{1,2} Another alternative is to use a conjugate gradient approach,¹ which avoids the use of a Hessian. The next point is given by:

$$\bar{x}_{i+1} = \bar{x}_i + \alpha \bar{s}_i; \bar{s}_i = -\bar{g}_i + [(\bar{g}_i - \bar{g}_{i-1})^t \bar{g}_i / \bar{g}_i^t \bar{g}_i] \bar{g}_{i-1} \quad (9)$$

where α is found by a line search that minimizes the energy.

At the end of each optimization step, there is a test for convergence. Checking the difference in the energy is not very sensitive, because the energy changes very little near the minimum. Examining the gradient (magnitude, root mean square, and/or maximum component) is a better convergence test. It may also be desirable to test the predicted displacements, since for fairly flat potential energy surfaces these can be large even if the gradient is small.

After an optimization has converged, one should test whether the structure is a true minimum. Since optimizations tend to retain the starting symmetry (because gradients belong to the totally symmetric representation), a lower-energy, lower-symmetry structure could exist nearby. This can be checked by calculating the full second derivative matrix or by calculating the vibrational frequencies (the Hessian used for the optimization is not sufficient since it normally contains no information about lower-symmetry distortions). For a minimum, all of the eigenvalues must be positive or all of the vibrational frequencies must be real.

3 METHODS FOR FINDING TRANSITION STRUCTURES

Quasi-Newton methods are quite reliable for minimizations. Similarly robust algorithms are not available for transition structures, since it is usually not possible to partition the problem in advance into a maximization in a fixed one-dimensional

space and a minimization in the remaining space. The direction of maximization normally must be determined during the course of the optimization. A variety of methods for finding transition structures exist and are discussed in a number of review articles^{2,7} (see *Transition Structure Optimization Techniques*).

One approach for optimizing a transition structure would be to transform the transition structure search into a minimization. Since the norm of the gradient is zero at the transition structure, one could minimize the gradient norm.⁸ However, the gradient norm surface has many more minima and maxima than the original surface, not just at the stationary points of the original surface but also at bumps and shoulders of the original surface. Thus a gradient norm optimization could easily converge to an undesired feature. The GDIIS method can have similar problems unless an appropriate transition vector is maintained in the Hessian. In the associated surface method⁹ a transition structure search is changed into a minimization by changing the sign of the lowest eigenvalue so that the Hessian is positive definite. If started within the quadratic region of the transition structure, this approach will converge. Under unfavorable circumstances, it may converge to some other low-lying transition structure on the surface, rather than to the desired transition structure.

A regular quasi-Newton optimization will converge to a transition structure provided that the initial guess is within the quadratic region and the Hessian has a suitable negative eigenvalue and eigenvector at each step in the optimization. However, the quadratic region around a transition structure may be much smaller than around a minimum. Consequently, the radius of convergence of a simple quasi-Newton algorithm may be much smaller, and special techniques need to be employed to get close to the transition structure.

The linear synchronous transit (LST) and quadratic synchronous transit (QST) approaches¹⁰ may be useful for getting closer to the transition structure. In the LST approach, the reaction path between reactants and products is approximated by a straight line (usually in distance matrix space or in internal coordinates) and a maximum is found along this line. Figure 3 shows some examples of LST and QST paths. The

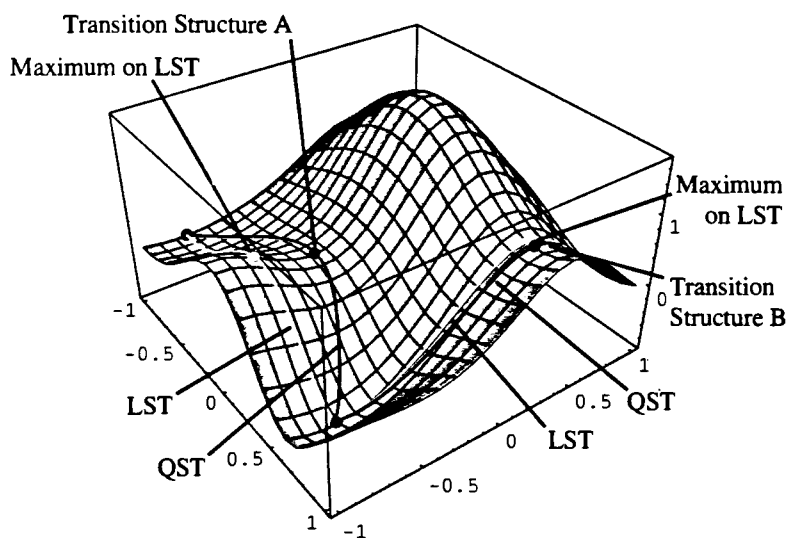


Figure 3 A model potential energy surface illustrating linear synchronous transit (LST) and quadratic synchronous transit (QST) paths (Reproduced from Ref. 2a, with permission from World Scientific Publishing)

maximum on the LST path is an upper bound to the transition structure energy, but may be some distance from the true transition structure if the actual reaction path is very curved. Under favorable circumstances, it may be close enough to the quadratic region of the transition structure for a quasi-Newton method to converge.

If a reaction is dominated by one coordinate (e.g., a torsion angle for a conformational change, or a bond length for a bond making/breaking reaction), then it may be possible to step along this coordinate while optimizing the remaining coordinates. This is termed coordinate driving. Under favorable circumstances, one can climb uphill toward the desired transition structure, but at the cost of numerous constrained optimizations. In unfavorable cases, one may end up climbing the wrong hill or walking up the wrong valley.

Alternative approaches that try to climb to the transition state from both sides simultaneously can be more successful.¹¹ A small step is taken from a point on the reactant side toward a point on the product side. Then the energy is minimized in a plane perpendicular to the step or such that the distance to the point on the product side is constant. A similar step is taken for the point on the product side. After several steps, when the two points are close enough to each other, the transition structure can be found by interpolation and quasi-Newton optimization.

With a suitable initial guess for the transition structure (if necessary, using one of the methods discussed above to get close to the transition structure), a modified quasi-Newton method can be used to optimize the transition structure. The flowchart is the same as Figure 2. A family of related procedures for quasi-Newton-like transition structure optimization go by names such as RFO, eigenvector following (EF), and walking up valleys.⁵ These methods take a controlled step toward a maximum along the lowest eigenvector and toward a minimum in the remainder of the space. This can be developed either in the trust radius approach (equation 5), or by the RFO method (equation 7), and is best done in terms of the eigenvectors and eigenvalues of the Hessian. If b_i and \bar{v}_i are the eigenvalues and eigenvectors of the Hessian, then the RFO or EF step toward a transition structure can be written analogously to equation (6):

$$\bar{x}_{i+1} = \bar{x}_i - \alpha \left(\frac{\bar{v}_1(\bar{v}_1^T \bar{g}_i)}{b_1 - \lambda_0} + \sum_{j \geq 2} \frac{\bar{v}_j(\bar{v}_j^T \bar{g}_i)}{b_j - \lambda} \right) \quad (10)$$

There are several ways of choosing λ_0 , λ , and α so that the optimization takes a step toward the transition structure but does not exceed the trust radius.⁵ However, the following conditions must be fulfilled: to step uphill in the shallowest ascent direction, $b_1 < \lambda_0 < b_2/2$; to step downhill in the remainder of the space, $\lambda \leq 0$ and $\lambda < b_2$.

The RFO and EF family of optimization methods can proceed toward the transition structure even when started outside the quadratic region. However, the Hessian must have a suitable lowest eigenvector that leads uphill to the appropriate transition structure. It is possible to follow an eigenvector other than the lowest one by rescaling the coordinate system so that the desired eigenvector is the lowest one. Note, however, that not all transition structures can be reached by EF from a minimum.

Since the Hessian for a transition structure search must have one negative eigenvalue, the BFGS method is unsuitable

since it is usually positive definite. The Bofill update¹² appears to be quite satisfactory and consists of a combination of the symmetric Powell and Murtagh-Sargent updates:

$$\Delta \mathbf{H}_{\text{SP}} = [(\Delta \bar{g} - \mathbf{H} \Delta \bar{x}) \Delta \bar{x}^T + \Delta \bar{x} (\Delta \bar{g} - \mathbf{H} \Delta \bar{x})^T] / \Delta \bar{x}^T \Delta \bar{x} - [(\Delta \bar{g} - \mathbf{H} \Delta \bar{x})^T \Delta \bar{x} \Delta \bar{x}^T] / (\Delta \bar{x}^T \Delta \bar{x})^2 \quad (11)$$

$$\Delta \mathbf{H}_{\text{MS}} = (\Delta \bar{g} - \mathbf{H} \Delta \bar{x}) (\Delta \bar{g} - \mathbf{H} \Delta \bar{x})^T / \Delta \bar{x}^T (\Delta \bar{g} - \mathbf{H} \Delta \bar{x}) \quad (12)$$

$$\Delta \mathbf{H}_{\text{Bofill}} = \phi \Delta \mathbf{H}_{\text{SP}} + (1 - \phi) \Delta \mathbf{H}_{\text{MS}}$$

$$\phi = 1 - [(\Delta \bar{x}^T (\Delta \bar{g} - \mathbf{H} \Delta \bar{x}))^2] / \Delta \bar{x}^T (\Delta \bar{g} - \mathbf{H} \Delta \bar{x})^2 \quad (13)$$

The line search for transition state optimization must be done more carefully than for minimization. If the step is predominantly along the transition vector, one needs to search for a maximum; if it is predominantly in the remaining space, one must search for a minimum. However, if the step has significant contributions from both parts, then the line search should be skipped, or it should look for a minimum in the gradient norm.

A difficulty with quasi-Newton methods for transition structure searching lies in choosing and maintaining a suitable direction for the transition vector. The LST or QST path can provide a guide for the optimization. Several gradient-based methods are available that alternate between maximizing along the path and minimizing perpendicular to the path.¹³ Another method uses a QST path to control an EF approach.¹⁴ Starting from a reactant-like structure and a product-like structure, a few steps are taken toward the maximum along the LST path. Then an EF method is used to converge to the transition structure. At each step the QST path through the current estimate of the transition structure and the reactant-like and product-like structures is used to select the appropriate eigenvector of the Hessian to be used for the EF method.

The steepest descent reaction path (see *Reaction Path Following*) is usually calculated after the transition structure has been located. There are some approaches that calculate both in the same procedure. These methods start with an approximate reaction path represented by a series of structures interpolated between reactants and products. The path is successively improved by a series of relaxation steps. Elber and Karplus^{15a} refined the path by minimizing the integral of the energy along the path under the constraint of equally spaced points:

$$S = \frac{1}{L} \int_{\bar{x}_0}^{\bar{x}_N} E(\bar{x}(s)) ds \approx \frac{1}{L} \sum_{i=1}^N \frac{E(\bar{x}_i) + E(\bar{x}_{i-1})}{2} |\bar{x}_i - \bar{x}_{i-1}| \quad (14)$$

Stacho and Ban^{15b} improve the path by taking a steepest descent step for each point and then redistributing the points to maintain equal spacing. Ayala and Schlegel^{15c} refine the path by optimizing the highest point to the transition structure and requiring the remaining points to satisfy the steepest descent reaction path equation. Early implementations were used with molecular mechanics methods and required a large number of steps;^{15a,b} recent methods are nearly as efficient as conventional transition structure and reaction path following calculations on the same reaction, and are suitable for electronic structure calculations.^{15c}

Once a transition structure optimization has converged, one must verify that it is a true transition structure and that it connects the desired reactants and products. The full Hessian

needs to be calculated at the same level of theory as the optimization. There must be one and only one negative eigenvalue for the structure to be a transition structure. Equivalently, there must be one and only one imaginary frequency. The eigenvector corresponding to the negative eigenvalue or the normal mode corresponding to the imaginary frequency should be inspected to confirm that motion along this coordinate connects reactants to products. A scan of the energies along an LST path connecting the transition structure to reactants and to products can be used to check if there are any additional transition structures along the reaction path. A better approach is to follow the steepest descent reaction path or intrinsic reaction coordinate from the transition structure to reactants and to products (see *Reaction Path Following*).

4 RELATED ARTICLES

Geometry Optimization: 2; Gradient Theory; Nonadiabatic Derivative Couplings; Reaction Path Following; Transition State Theory; Transition Structure Optimization Techniques.

5 REFERENCES

- (a) R. Fletcher, 'Practical Methods of Optimization', Wiley, Chichester, 1981; (b) P. E. Gill, W. Murray, and M. H. Wright, 'Practical Optimization', Academic Press, New York, 1981; (c) M. J. D. Powell, (ed.), 'Non-linear Optimization, 1981', Academic Press, New York, 1982; (d) J. E. Dennis and R. B. Schnabel 'Numerical Methods for Unconstrained Optimization and Non-linear Equations', Prentice-Hall, Englewood Cliffs, NJ, 1983; (e) L. E. Scales, 'Introduction to Non-linear Optimization', Macmillan, Basingstoke, 1985.
- (a) H. B. Schlegel, in 'Modern Electronic Structure Theory', ed. D. R. Yarkony, World Scientific, Singapore, 1995, pp. 459-500; (b) T. Schlick, *Rev. Comput. Chem.*, 1992, **3**, 1-71; (c) H. B. Schlegel, *Adv. Chem. Phys.*, 1987, **67**, 249-286; (d) D. Heidrich, W. Kliesch, and W. Quapp, 'Properties of Chemically Interesting Potential Energy Surfaces', Springer-Verlag, Berlin, 1991.
- (a) A. Törn and A. Zilinskas, 'Global Optimization', Springer-Verlag, Berlin, 1989; (b) R. Horst and H. Tuy 'Global Optimization', Springer-Verlag, Berlin, 1990; (c) C. A. Floudas and P. M. Pardalos (eds.), 'Recent Advances in Global Optimization', Princeton University Press, Princeton, NJ, 1992; (d) R. Horst and P. M. Pardalos, 'Handbook of Global Optimization: Nonconvex Optimization and Its Applications', Kluwer, Dordrecht, 1994; (e) C. A. Floudas and P. M. Pardalos, 'State of the Art in Global Optimization: Computer Methods and Application', Kluwer, Dordrecht, 1996.
- (a) P. Pulay, G. Fogarasi, F. Pang, and J. E. Boggs *J. Am. Chem. Soc.*, 1979, **101**, 2550-2560; (b) P. Pulay and G. Fogarasi, *J. Chem. Phys.*, 1992, **96**, 2856-2860; (c) G. Fogarasi, X. Zhou, P. Taylor, and P. Pulay, *J. Am. Chem. Soc.*, 1992, **114**, 8191-8201; (d) J. Baker, *J. Comput. Chem.*, 1993, **14**, 1085-1100; (e) C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch, *J. Comput. Chem.*, 1996, **17**, 49-56.
- (a) A. Banerjee, N. Adams, and J. Simons, *J. Phys. Chem.*, 1985, **89**, 52-57; (b) J. Simons and J. Nichols, *Int. J. Quantum Chem., Quantum Chem. Symp.*, 1990, **24**, 263-276 and references therein; (c) J. Baker, *J. Comput. Chem.*, 1986, **7**, 385-395.
- P. Csaszar and P. Pulay, *J. Mol. Struct.*, 1984, **114**, 31-34.
- (a) S. Bell and J. S. Crighton, *J. Chem. Phys.*, 1984, **80**, 2464-2475; (b) J. D. Head, B. Weiner, and M. C. Zerner, *Int. J. Quantum Chem.*, 1985, **33**, 177-186; (c) D. A. Liotard, *Int. J. Quantum Chem.*, 1992, **44**, 723-741.
- (a) J. W. McIver, Jr., and A. Komornicki, *J. Am. Chem. Soc.*, 1972, **94**, 2625-2633; (b) A. Komornicki, K. Ishida, K. Morokuma, R. Ditchfield, and M. Conrad, *Chem. Phys. Lett.*, 1977, **45**, 595-602; (c) D. Poppinger, *Chem. Phys. Lett.*, 1975, **35**, 550-554.
- T. Helgaker, *Chem. Phys. Lett.*, 1991, **182**, 503-510.
- T. A. Halgren and W. N. Lipscomb, *Chem. Phys. Lett.*, 1977, **49**, 225-232.
- (a) K. Müller and L. D. Brown, *Theor. Chim. Acta*, 1979, **53**, 75-93; (b) M. J. S. Dewar, E. F. Healy, and J. J. P. Stewart, *J. Chem. Soc., Faraday Trans. 2*, 1984, **80**, 227-233; (c) S. Fischer and M. Karplus, *Chem. Phys. Lett.*, 1992, **194**, 252-261; (d) C. Cardenas-Lailhacar and M. C. Zerner, *Int. J. Quantum Chem.*, 1995, **55**, 429-439.
- J. M. Bofill, *J. Comput. Chem.*, 1994, **15**, 1-11.
- S. Bell and J. S. Crighton, *J. Chem. Phys.*, 1984, **80**, 2464-2475.
- C. Peng and H. B. Schlegel, *Isr. J. Chem.*, 1994, **33**, 449-454.
- (a) R. Elber and M. Karplus, *Chem. Phys. Lett.*, 1987, **139**, 375-380; (b) L. L. Sacho and M. I. Bán, *Theor. Chim. Acta*, 1992, **83**, 433-440; (c) P. Ayala and H. B. Schlegel, *J. Chem. Phys.*, 1997, **107**, 375-384.

Geometry Optimization: 2

Tamar Schlick

The Howard Hughes Medical Institute and New York University, NY, USA

1	Introduction	1143
2	Basic Definitions of Optimization Problems	1143
3	Optimization Fundamentals	1144
4	Basic Algorithmic Components	1146
5	Newton's Method	1148
6	Effective Large-scale Minimization Algorithms	1150
7	Available Software and Practical Recommendations	1153
8	Looking Ahead	1155
9	Appendix A: The Quadratic Convergence of Newton's Method	1155
10	Appendix B: The Attainable Accuracy of Newton's Method	1156
11	Related Articles	1156
12	References	1156

Abbreviations

CG = conjugate gradient; CPU = central processing units; QN = quasi Newton; SD = steepest descent; TN = truncated Newton.

Glossary

Convergence order, p

The largest number such that a finite limit β (the convergence ratio) exists for a sequence $\{x_r\}$ where

$$0 \leq \lim_{k \rightarrow \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|^p} = \beta < \infty$$