In this model, reaction 3 is faster than reaction 4, if atomic sulfur accumulates. Atomic sulfur has not been found in colloidal or powder suspension systems containing a sacrificial electron donor, so that there seems to be a direct involvement of the host Nafion. Finally, since the S/Cd ratio for the irradiated cubic CdS sample approaches that for CdS powder after a short period of Ar^+ bombardment, the atomic sulfur layer appears to be, at most, a few monolayers thick.

As indicated in Table II, the surface of the cubic CdS film soaked in Na_2S solution without irradiation is covered with excess S^{2-} ion of Na_2S. Surface sulfate species is also present. The data indicate that the S^{2-} ion of Na_2S or H_2S is strongly adsorbed on the CdS–Nafion surface and is oxidized to sulfate (Figures 4 and 6) without irradiation.

5. Conclusions

On the basis of the results reported here, we draw the following conclusions:

(1) The concentration of *surface* cation exchange sites in Nafion is small.

(2) The small CdS particles at the surface of the hexagonal CdS films are subject to dissolution in boiling water, whereas the large CdS particles at the surface of the cubic CdS films are stable in boiling water.

(3) After extensive washing, the surface of films containing cubic CdS remains dominated by sulfide ions of CdS, whereas the surface of films containing the hexagonal form are altered, leaving surface sulfate associated with the Nafion and Cd^{2+} associated with cation exchange sites.

(4) Adsorbed sulfide ions on Nafion and CdS–Nafion are oxidized to sulfate ions at 300 K in the presence of oxygen.

(5) The gray-blue deposit formed on cubic CdS–Nafion surfaces under irradiation is identified as atomic sulfur.

Acknowledgment. We acknowledge financial support of this work by the Gas Research Institute (Contract No. 5982-260-0756). National Science Foundation support of the X-ray photoelectron spectrometer by an equipment grant, CHE 8201179, is also acknowledged.

Registry No. CdS, 1306-23-6; Nafion, 39464-59-0.

Search for Stationary Points on Surfaces

Ajit Banerjee, Noah Adams, Jack Simons,*

Department of Chemistry, The University of Utah, Salt Lake City, Utah 84112

and Ron Shepard

Theoretical Chemistry Group, Argonne National Laboratory, Argonne, Illinois 60439 *(Received: July 9, 1984)*

Algorithms for finding local minima, maxima, and saddle points on surfaces, starting from an arbitrary point, are presented. These algorithms are based on making a local approximation to the surface in the form of a rational function constructed from the local first and second derivatives of the surface. All parameters of these algorithms required for stepping across the surface are determined in nonarbitrary ways. The convergence of these procedures to the desired stationary point is shown to be quadratic. Applications for stationary-point searches on two model surfaces are also given for illustrative purposes.

Introduction

Finding stationary points (minima, maxima, and saddle points) on energy surfaces is important in chemical physics because they correspond to equilibrium geometries and transition states and because the classical equations of motion connecting such points can be used to describe the reaction dynamics. For a surface $E(x_1,...,x_n)$ depending on independent variables $\{x_i\}$, a stationary point is characterized by $\partial E/\partial x_i = 0$, that is, by vanishing slopes for all variables. In addition, a minimum corresponds to a point having positive curvatures $\partial^2 E/\partial x_i^2 > 0$, for all variables and a maximum has $\partial^2 E/\partial x_i^2 < 0$. For a saddle point of order μ, there exist μ variables relative to which the curvature is negative; it is positive relative to the remaining $(n - \mu)$ variables. In molecular quantum chemistry the ab initio determination of first and second derivatives of the energy with respect to geometric variables, though laborious, is within reach.[1] An algorithm that systematically and efficiently locates stationary points starting from an arbitrary point and using local first- and second-derivative information is given in this paper.

There exists extensive literature in numerical analysis regarding the search for extrema on multidimensional surfaces. References 2–6 describe techniques for minima and saddle point searches. The highlights of the algorithm presented here are that all optimization parameters are calculated in an ab initio manner and the procedure is quadratically convergent with the stability characteristics of a saddle-point search being the same as those of a local-minimum search.

General Discussion

For a surface of n independent variables, the Taylor expansion allows one to explore the neighborhood of a point x_0

$$E(\mathbf{x}) = E_0 + \mathbf{g}^+\mathbf{x} + \tfrac{1}{2}\mathbf{x}^+\boldsymbol{H}\mathbf{x} + ... \quad (1)$$

Here the step vector $\mathbf{x} = (x_1, ..., x_n)$ gives the displacement away from $\mathbf{x}_0$, the gradient vector $\mathbf{g}$ and hessian matrix $\boldsymbol{H}$ contain elements of first and second derivatives of E with respect to the (x_i) calculated at $\mathbf{x}_0$. The Newton–Raphson (NR) procedure allows one to step toward a stationary point of E, near $\mathbf{x}_0$ by approximating the expansion of eq 1 in the neighborhood of $\mathbf{x}_0$ quadratically

$$\epsilon = E(\mathbf{x}) - E_0 = \mathbf{g}^+\mathbf{x} + \tfrac{1}{2}\mathbf{x}^+\boldsymbol{H}\mathbf{x} \quad (2)$$

and imposing the stationary requirement $\partial\epsilon/\partial x_i = 0$, giving

$$\boldsymbol{H}\mathbf{x} + \mathbf{g} = 0 \quad (3)$$

To facilitate the analysis of the NR algorithm and those developed in this paper, without any loss of generality, it is useful to transform to new displacement variables in which the Hessian matrix is

(1) Banerjee, A.; Jensen, J.; Simons, J.; Shepard, R. *Chem. Phys.* **1984**, *87*, 203. Jørgensen, P.; Simons, J. *J. Chem. Phys.* **1983**, *79*, 334.
(2) Fletcher, R. "Practical Methods of Optimization"; Wiley: New York, 1980 and references therein.
(3) Simons, J.; Jørgensen, P.; Taylor, H.; Ozment, J. *J. Phys. Chem.* **1983**, *87*, 2745.
(4) Cerjan, C. J.; Miller, W. H. *J. Chem. Phys.* **1981**, *75*, 2800.
(5) Bell, S.; Crighton, J. C. *J. Chem. Phys.* **1984**, *80*, 2464.
(6) Lootsma, F. A., Ed. "Numerical Methods for Non-linear Optimization"; Academic Press: New York, 1972.

0022-3654/85/2089-0052$01.50/0

diagonalized (by a unitary matrix U):

$$H \rightarrow h = U^{+} H U = \begin{pmatrix} h_1 & & 0 \\ & \ddots & \\ 0 & & h_n \end{pmatrix}, \quad \mathbf{x} \rightarrow U^{+}\mathbf{x}, \quad \mathbf{g} \rightarrow U^{+}\mathbf{g} \tag{4}$$

Such variables are called the local principal modes or axes. In this representation the NR step along each principal axis is, from eq 3

$$x_i = -g_i/h_i \quad i = 1, \ldots, n \tag{5}$$

The stationary points, in this representation, have $g_i = 0$ for all i, and Hessian eigenvalues ordered according to $0 \leq h_1 \leq h_2 \ldots \leq h_n$ for a minimum, $h_1 \leq h_2 \ldots \leq h_n \leq 0$ for a maximum, and $h_1 \leq \ldots \leq h_\mu \leq 0 \leq h_{\mu+1} \ldots \leq h_n$ for a saddle point of order μ.

The above straightforward NR procedure, in the neighborhood of a minimum ($h_i > 0$) takes a step in the direction *opposite* to the gradient along each mode, that is, it steps toward the minimum. Similarly, it steps *along* the gradient toward the maximum. Generally, the NR algorithm steps toward the nearest stationary point on the surface by following opposite ($h_i > 0$) and along ($h_i < 0$) the various gradient directions.

However, it is often the case that, having located a stationary point, one desires to step *away* from its neighborhood in search of another stationary point perhaps with different character. There is a particular modification of the above NR stepping algorithm, which forms an essential basis for this paper, which can be used to guide the NR-like steps away from a nearby stationary point in search of other critical points. It gives the step lengths along each of the local modes as

$$x_i = -g_i/(h_i - \lambda) \tag{6}$$

where λ is an appropriately chosen shift parameter. Depending upon the value of λ, the sign of each $(h_i - \lambda)$ will be positive or negative, and hence the direction of the step x_i will be opposite or toward the direction of the gradient. In the following discussion we further explore this concept by examining which functional forms of E produce a stepping algorithm of the form in eq 6. The early work of Cerjan and Miller resulted in this type of step formula, but their prescription for choosing the parameter λ differed from that proposed here. In this paper, we therefore focus on how λ can be chosen in some optimal sense, and whether more than one such λ variable can be of additional use.

The Rational Function Optimization (RFO) Approaches

Following methods used for the optimization of the parameters of the electronic wave function for ground or excited states,[7,8] we can rewrite the local representation of the surface of eq 2 in a form which utilizes the same local derivative data as contained in the Taylor expansion but which is of the rational function type:

$$\epsilon = E(\mathbf{x}) - E_0 = \frac{\mathbf{g}^{+}\mathbf{x} + {}^1/_2\mathbf{x}^{+}\boldsymbol{H}\mathbf{x}}{1 + \mathbf{x}^{+}\boldsymbol{S}\mathbf{x}} = \frac{{}^1/_2(\mathbf{x}^{+}1)\begin{pmatrix} \boldsymbol{H} & \mathbf{g} \\ \mathbf{g}^{+} & 0 \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix}}{(\mathbf{x}^{+}1)\begin{pmatrix} \boldsymbol{S} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix}} \tag{7}$$

where the elements of the symmetric $\boldsymbol{S}$ matrix are to be specified as described below. This form can be interpreted as a [2/2] Padé approximant to ϵ. Taking the first and second derivatives of this function at $\mathbf{x}_0$, one sees that $(\partial\epsilon/\partial x_i)_0 = g_i$ and $(\partial^2\epsilon/\partial x_i\,\partial x_j)_0 = H_{ij}$, which are identical with those obtained in eq 2. This results because the denominator of eq 7 expanded as $(1 + \mathbf{x}^{+}\boldsymbol{S}\mathbf{x})^{-1}$ contributes to ϵ only through third and higher orders in $\mathbf{x}$. In contrast to eq 2, ϵ and its gradients possess finite $x \rightarrow \pm\infty$ asymptotes, which are determined by the Hessian elements and the scaling matrix $\boldsymbol{S}$.

The stationary condition requirement $\partial\epsilon/\partial x_i = 0$ applied to the model surface of eq 7 then yields an eigenvalue equation of dimension $(n + 1)$:

$$\begin{pmatrix} \boldsymbol{H} & \mathbf{g} \\ \mathbf{g}^{+} & 0 \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \lambda\begin{pmatrix} \boldsymbol{S} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} \tag{8}$$

where the eigenvalue $\lambda = 2\epsilon$ gives twice the change in ϵ accompanying the step along $\mathbf{x}$, which is determined from an intermediately normalized eigenvector (i.e., an eigenvector having the value unity as its last entry) of this equation.

To analyze further and interpret the eigensolutions of this RFO algorithm and to motivate a specific choice for the scaling matrix $\boldsymbol{S}$, let us rewrite eq 8 in a partitioned form as the following two equations:

$$(\boldsymbol{H} - \lambda\boldsymbol{S})\mathbf{x} + \mathbf{g} = \mathbf{0} \tag{9a}$$

$$\mathbf{g}^{+}\mathbf{x} = \lambda \tag{9b}$$

The matrix $\boldsymbol{S}$ clearly has units since $\mathbf{x}^{+}\boldsymbol{S}\mathbf{x}$ must be dimensionless. In principle, many choices could be made for the $\boldsymbol{S}$ matrix given information about the shape of the energy surface near the boundary region where the quadratic approximation $\mathbf{g}^{+}\mathbf{x} + {}^1/_2\cdot\mathbf{x}^{+}\boldsymbol{H}\mathbf{x}$ fails to be reasonable. In our opinion, there are only two reasonable choices to make for $\boldsymbol{S}$ given the limited data which we assume to be available in developing this surface walking algorithm; one could choose to scale all Cartesian directions uniformly $\boldsymbol{S} = a^{-2}\boldsymbol{I}$ or to make a more direction-specific scaling by examining the behavior of the algorithm along the eigenmodes of the Hessian matrix (see below). The uniform-scaling choice allows eq 8 to be rewritten as

$$\begin{pmatrix} a^2\boldsymbol{H} & a\mathbf{g} \\ a\mathbf{g}^{+} & 0 \end{pmatrix}\begin{pmatrix} \mathbf{x}/a \\ 1 \end{pmatrix} = 2\epsilon\begin{pmatrix} \mathbf{x}/a \\ 1 \end{pmatrix} \tag{9c}$$

in which the eigenvalue λ remains 2ϵ, but the eigenvector contains not the $\mathbf{x}$ step-length components but scaled components $\mathbf{x}/a$; the gradient or force elements are also scaled to $a\mathbf{g}$ and the Hessian is scaled to $a^2\boldsymbol{H}$. This same scaled eigenvalue problem (eq 9c) could, alternatively, have been obtained by formulating the local energy surface approximation (eq 7) in terms of dimensionless coordinates $\mathbf{y} = \mathbf{x}/a$.

An alternative, and slightly more general, choice of $\boldsymbol{S}$ can be made by examining eq 9a and 9b within the basis which diagonalizes the Hessian matrix. If $\boldsymbol{S}$ is taken to be a diagonal matrix in this basis with diagonal elements a_i^{-2}, then eq 8 reads

$$\begin{pmatrix} h_1 & 0 & 0 \ldots & & g_1 \\ 0 & h_2 & 0 \ldots & & \\ \vdots & & & h_n & g_n \\ & & & g_n & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ 1 \end{pmatrix} =$$

$$2\epsilon\begin{pmatrix} a_1^{-2} & 0 & 0 & \ldots & 0 \\ 0 & a_2^{-2} & 0 & \ldots & 0 \\ \vdots & & a_n^{-2} & & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ 1 \end{pmatrix} \tag{10}$$

In our opinion, the simple eigenvalue structure of eq 9c is to be preferred over eq 10, especially since one simply does not usually have enough information available to permit the n potentially different scale parameters $\{a_i\}$ to be chosen. Therefore, we shall restrict our attention for the duration of this paper to examining the performance of eq 8, 9a, and 9b, to which even eq 10 reduces if all of the $\{a_i\}$ are chosen to be identical (i.e., if $a_i = a$ for all i). In summary then, we assume that all of the Cartesian (i.e., length) coordinates $\{x_i\}$ have been scaled by a common amount a (i.e., $\mathbf{x} \leftarrow \mathbf{x}/a$) and the gradient and Hessian elements calculated in terms of these scaled coordinates. Hence, in the following equations the scale parameter is absent, having been absorbed into the definition of the dimensionless coordinates (i.e., $\boldsymbol{S}$ is taken to be the unit matrix).

Returning now to an analysis of the implications of using eq 9 to carry out surface walks, let us move to the representation in which the Hessian $\boldsymbol{H}$ is diagonal. In this representation, the step lengths along each mode are

$$x_i = g_i/(\lambda - h_i) \tag{11}$$

(7) Banerjee, A.; Grein, F. *Int. J. Quantum Chem.* **1976**, *10*, 123.
(8) Shepard, R.; Shavitt, I.; Simons, J. *J. Chem. Phys.* **1982**, *76*, 543.
(9) Löwdin, P. O. In "Proceedings of an Advanced Seminar on Perturbation Theory and Its Applications in Quantum Mechanics, Oct 1965, Madison, WI"; Wilcox, C. H., Ed.; Wiley: New York, 1966.

which is indeed of the form introduced in eq 6. Most importantly, now we have a prescription for choosing the so-called Hessian eigenvalue shift parameter; $\lambda = 2\epsilon$ is to be calculated as an eigenvalue of the RFO matrix of eq 8 with $\boldsymbol{S}$ now equal to the identity matrix $\boldsymbol{I}$ since we are dealing with scaled coordinates. Equation 11, when substituted into eq 7 and eq 9b, gives an expression for the change in the function ϵ and an equation which can be used to evaluate λ:

$$\epsilon = 1/z \sum_i g_i^2(\lambda - h_i/2)/(\lambda - h_i)^2 \qquad (12)$$

$$\sum_i g_i^2/(\lambda - h_i) = \lambda \qquad (13)$$

where $z = (1 + \sum_i x_i^2)$ is always a positive quantity. Considering eq 11, 12, and 13 as parametric forms in λ, one observes that the function $E(\mathbf{x})$ is predicted to go up or down (ϵ positive or negative) along the ith mode depending on whether $(\lambda - h_i/2)$ is positive or negative. The sign of $(\lambda - h_i)$ from eq 11 will determine whether the step along the ith mode is toward or opposite the gradient. The parameter values λ for which the left-hand side (lhs) and right-hand side (rhs) of eq 13 are equal correspond to the eigenvalues of eq 8.

To appreciate the behavior of the solutions, it is instructive to plot the two sides of eq 13 as functions of λ, as is shown in Figure 1a. The vertical asymptotes correspond to the eigenvalues of the Hessian $h_1, h_2, \ldots, h_n$. The lhs tends to $+\infty$ (or $-\infty$) as λ approaches an eigenvalue h_i from the right (or left). For large positive (negative) values of λ, the lhs is positive (negative) and tends to zero. The lhs thus has $n + 1$ branches. The rhs, on the other hand, is a straight line through the origin with unit slope. The points of intersection, $n + 1$ of them, correspond to the eigenvalues of eq 8. It is clear from Figure 1a that eigenvalues of eq 8 bracket the eigenvalues of the Hessian, $\lambda_i \leq h_i \leq \lambda_{i+1}$. This result is known as a separation or bracketing theorem.[9]

At convergence to a minimum-energy point, all $g_i = 0$, which gives $x_i = 0$. The corresponding lowest eigenvalue λ of the RFO matrix is also zero and the other n eigenvalues are those of the Hessian at the minimum point. For a saddle point of order μ, the zero eigenvalue separates the μ negative and $(n - \mu)$ positive eigenvalue of the Hessian. In particular, near a local minimum, the corresponding lowest eigenvalue λ_1 is negative until at convergence, it is equal to zero.

Even though the above analysis is presented in the diagonal-Hessian representation, application of the method does not require that $\boldsymbol{H}$ be diagonal. One simply constructs the RFO matrix of eq 8 at $\mathbf{x}_0$, calculates the appropriate eigenvector (e.g., corresponding to the lowest eigenvalue for a minimum search), and then obtains the new starting point as $\mathbf{x}_0 = \mathbf{x}_0 + \mathbf{x}$. This process is continued until convergence is obtained. This RFO procedure can be shown to converge quadratically.[7,8] This may be confirmed by writing the derivative of the energy near the stationary point as

$$\partial E/\partial \boldsymbol{x} = \mathbf{g} + \boldsymbol{H}\mathbf{x} + O(\mathbf{x}^2) \qquad (14)$$

where $O(\mathbf{x}) = O(\mathbf{g}) < 1$. Therefore writing the RFO eq 9 as $\mathbf{x} = -\boldsymbol{H}^{-1}\mathbf{g} + O(\lambda\mathbf{x})$, and using $\lambda = \mathbf{g}^+\mathbf{x}$, $O(\lambda) = O(\mathbf{x}^2)$ one has $\partial E/\partial \boldsymbol{x} = O(\mathbf{x}^2) + O(\mathbf{x}^3)$. The gradient, and therefore the energy change in each iteration, decreases as the square of the value in the previous iteration and hence the algorithm is quadratically convergent in the neighborhood of a solution. Furthermore, scaling of the eigenvector by a factor α used to calculate mode displacements preserves quadratic convergence provided that $(1 - \alpha) = O(\mathbf{x}^2)$. Note that the intermediate normalization in eq 8 corresponds to $\alpha = (1 + \mathbf{x}^+\mathbf{x})^{-1/2}$.

Search for Local Minimum. Let us further analyze the behavior of the above algorithm during the stepwise search for a local minimum. The specific cases depicted in Figure 1, b–d, correspond to locations $\mathbf{x}_0$ on the surface whose local structure (characterized via the signs of the hessian eigenvalues) may or may not be consistent with that of the region of the desired stationary point. Figure 1b shows the case when one eigenvalue of the Hessian is positive and the other negative, which would be

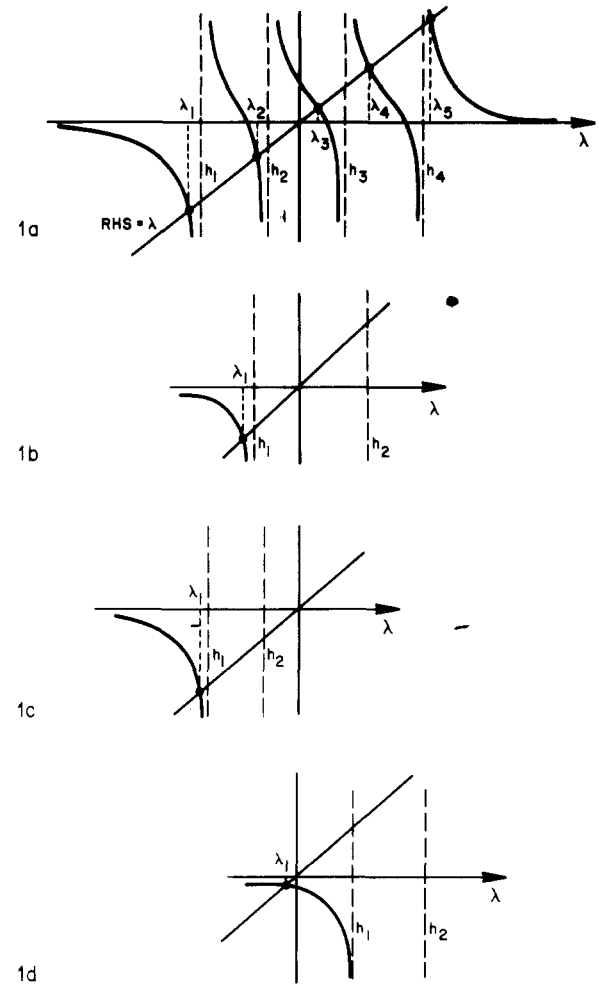

Figure 1. Schematic plots of the left- and right-hand sides of eq 13 as a function of λ. 1a shows all of the roots and 1b, 1c, and 1d correspond to a minimum search.

the case if the starting point $\mathbf{x}_0$ were near a saddle point. As discussed earlier, λ_1 is always negative so that the function value ϵ always decreases. In addition, since in this case $(\lambda_1 - h_1/2)$ and $(\lambda_1 - h_2/2)$ are both negative, components of the change in ϵ along each of the principal modes also decrease (including the mode with negative curvature). This is, of course, what one desires in searching for a local minimum. Figure 1c shows a case when two Hessian eigenvalues are negative (near a maximum) and Figure 1d depicts a case when the structure of the Hessian is consistent with a minimum. For all of these situations the above analysis holds; the function E and all its components are always lowered when $\lambda = \lambda_1$ is used, thereby assuring minimization of the ϵ function. Search for a local maximum follows analogous lines. Here, the largest eigenvalue, λ_{n+1} (for which the function value always increases), is chosen to define $\mathbf{x}$.

Search for Saddle Points. The RFO algorithm introduced in the last section can also be used for saddle-point searches. For example, to find a first-order saddle point, one would choose the second eigenvalue λ_2 and its eigenvector to determine the changes in the ϵ function and the step length and direction $\mathbf{x}$. However, let us first examine slight modifications in the algorithm, which are aimed at attempting to increase its stability for saddle-point searches. Levenberg and Marquardt have shown[10] that, for step-constrained extrema searching algorithms, only *one* parameter λ is required for the calculation of the optimum steps. For a saddle-point search of μ-th order, the above RFO algorithm of course gives only one λ (namely, $\lambda_{\mu+1}$). However, because of the bracketing property, the fact that $h_\mu \leq \lambda_{\mu+1} \leq h_{\mu+1}$ (see Figure 1a) suggests a possible separation of the problem into μ principal

(10) Davies, M.; Whitting, I. J. ref 6, p 191.

modes to be maximized and $(n - \mu)$ principal modes to be minimized. Such a thought raises the question of whether it is it possible to find two curvature shift parameters, λ_p and λ_n, one for modes relative to which the ϵ function is to be maximized and the other for which it is minimized. Such an approach can be developed by solving separately two partitioned problems: maximization of ϵ along a chosen set of μ-principal modes and then minimization along the remaining n - μ principal modes.

This division into two separate optimization problems can be accomplished by constructing a unitary matrix $\boldsymbol{U} = [\mathbf{u}_1, ..., \mathbf{u}_\mu, v_1, ..., v_{n-\mu}]$, where a set of normalized linearly independent column vectors $\mathbf{v}_i$ have been constructed to be orthogonal to the $\mathbf{u}_k$, which are eigenfunctions of the Hessian, $\boldsymbol{H}\,\mathbf{u}_k = \lambda_k \boldsymbol{u}_k$. Then $\boldsymbol{U}^+\boldsymbol{H}\boldsymbol{U}$ has the desired block diagonal form in which the elements of the diagonal blocks are $\mathbf{u}_k{}^+\boldsymbol{H}\mathbf{u}_l = \lambda_k\delta_{kl}$, and $\mathbf{v}_i{}^+\boldsymbol{H}\mathbf{v}_j$, and the elements of the off-diagonal block vanish $\mathbf{v}_i{}^+\boldsymbol{H}\mathbf{u}_k = \lambda_k v_i{}^+ u_k = 0$. Thus in this new basis the μ and $n - \mu$ dimensional eigenspaces of $\boldsymbol{H}$ are noninteracting. With the desired separation achieved, the partitioned RFO (P-RFO) algorithm yields the following two analogs of eq 8:

$$\left(\begin{array}{ccc|c} H_1 & & 0 & g_1 \\ & \ddots & & \vdots \\ 0 & & H_\mu & g_\mu \\ \hline g_1 & \cdots & g_\mu & 0 \end{array}\right)\left(\begin{array}{c} x_1 \\ \vdots \\ x_\mu \\ \hline 1 \end{array}\right) = \lambda_p \left(\begin{array}{c} x_1 \\ \vdots \\ x_\mu \\ \hline 1 \end{array}\right) \tag{15}$$

$$\left(\begin{array}{ccc|c} H_{\mu+1} & \cdots & H_{\mu+1,n} & g_{\mu+1} \\ \vdots & \ddots & \vdots & \vdots \\ H_{n,\mu+1} & \cdots & H_n & g_n \\ \hline g_{\mu+1} & \cdots & g_n & 0 \end{array}\right)\left(\begin{array}{c} x_{\mu+1} \\ \vdots \\ x_n \\ \hline 1 \end{array}\right) = \lambda_n \left(\begin{array}{c} x_{\mu+1} \\ \vdots \\ x_n \\ \hline 1 \end{array}\right) \tag{16}$$

when expressed in the $\{\mathbf{u}_k, \mathbf{v}_i\}$ basis. Here λ_p and λ_n are the highest and lowest eigenvalues of eq 15 and 16, respectively, λ_p is always positive, and λ_n is always negative and both approach zero near convergence.

For the application the above P-RFO algorithm to a search for a saddle point of order μ, one must decide *which* of the μ-principal modes to be maximized. For $\mu = 1$ this means that one must decide which local mode along which to maximize ϵ. This is not an easy question to answer. To appreciate the difficulty, imagine a walk from a region where the structure of the Hessian is not consistent with the desired saddle point. If the Hessian eigenvalues are smoothly varying functions along the step, the lowest μ-eigenvalues of $\boldsymbol{H}$ would need to undergo the least change to evolve into a Hessian of the correct structure. Hence, the principal modes corresponding to the μ smallest eigenvalues of $\boldsymbol{H}$ should be chosen for maximization and the $n - \mu$ remaining modes for minimization to search for a closest (in curvature change) saddle point. This assumption is consistent with the RFO algorithm based on the bracketing theorem (shown pictorially in Figure 1a) where the $(\mu + 1)$th eigenvalue of the RFO matrix always falls between the μth and $(\mu + 1)$th eigenvalues of $\boldsymbol{H}$ (arranged in ascending order). Unfortunately, however, it is not always true that choosing the μ *lowest* modes to be maximized will result in success. In general, it may be necessary to explore *all* possible choices of μ modes out of n total degrees of freedom for maximization to locate the desired saddle point.

The above P-RFO algorithm is based on extremization of two new functions using the RFO algorithm. Therefore it is quite stable and also converges quadratically. Unlike the original RFO algorithm, its application requires transformation to a basis in which the Hessian is at least block diagonal.

Comparison of the RFO and P-RFO Algorithms. Let us now analyze the comparative step-by-step behavior of the two algorithms (RFO, P-RFO) which have been proposed for use in searches for saddle points. For simplicity, in a–c of Figure 2 are shown representatives for the cases with $\mu = 1$. Figure 2a shows the behavior of the two algorithms when the structure of $\boldsymbol{H}$ is consistent with the desired first-order ($\mu = 1$) saddle point; $\boldsymbol{H}$ has one positive and one negative eigenvalue. The RFO algorithm produces a second root λ_2 which is positive (it can also be negative), because the starting point $\mathbf{x}_0$ was below the desired saddle point hence the function value must increase. As a result, $(\lambda_2 - h_1/2)$ is positive while $(\lambda_2 - h_2/2)$ is negative, so the component of the ϵ function along one-mode increases, while it decreases along the second mode, as desired. However, it is conceivable that $(\lambda_2 - h_2/2)$ could have also been positive. In contrast, the P-RFO produces $(\lambda_p - h_1/2)$ that is always positive and $(\lambda_n - h_2/2)$ that is always negative. Figure 2b contrasts the performances of the RFO algorithm and its variant at a point which is not in the neighborhood of the desired saddle point ($\mu = 1$). Here, both eigenvalues of $\boldsymbol{H}$ are positive (e.g., at a point $\mathbf{x}_0$ near a minimum

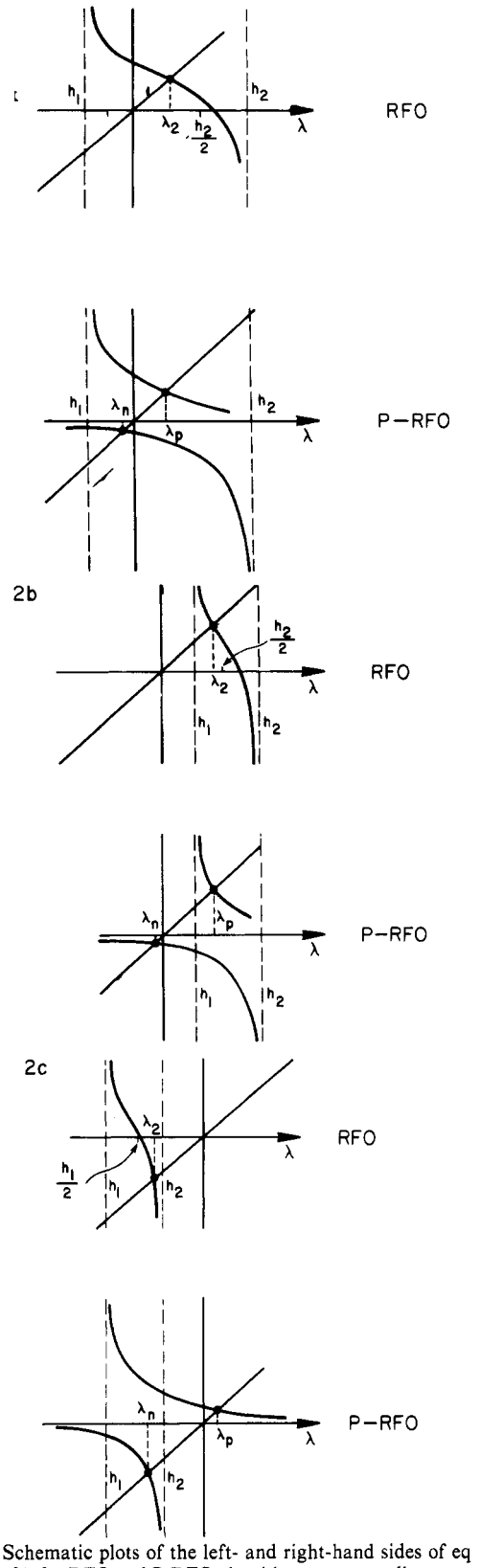

Figure 2. Schematic plots of the left- and right-hand sides of eq 13 as a function of λ for RFO and P-RFO algorithms corresponding to a search for a saddle point of first order.

with the desired saddle point above $\mathbf{x}_0$). The RFO algorithm gives λ_2 and $(\lambda_2 - h_1/2)$ always positive while $(\lambda_2 - h_2/2)$ can be negative (usual case) or positive. Thus the total function and its component along the first mode always increase. Pictorially, the RFO algorithm for such cases may step down toward (usual case) or away from the "valley floor" along the second mode while always going up the valley along the first mode in a manner which assures that the total function always increases. In contrast (and by construction), the P-RFO algorithm has $(\lambda_p - h_1/2)$ and $(\lambda_n - h_2/2)$ always positive and negative, respectively. That is, it always steps down toward the valley floor along the second mode while going up the valley along the first mode. However, it does *not* assure that the cumulative effect will result in an increase of the ϵ function value. Figure 2c shows a similar case when both eigenvalues of $\boldsymbol{H}$ are negative (near a maximum), where the same analysis applies and the usual step directions are up toward a ridge top along the 1-mode while going down the ridge in the 2-mode in search for a saddle point.

Let us now examine the relative magnitudes of the step lengths x_i along each of the principal directions.

Overstepping. So far we have assumed that, starting from a point $\mathbf{x}_0$ (and based on local approximations to the true surface of the type in eq 7), the properties of the new point $\mathbf{x}_1$ (gradient and Hessian) predicted by the stepping algorithms are consistent with those of the true surface within some tolerance. Overstepping into a region where the Hessian structure is not that which characterizes the desired stationary point is entirely possible. Unfortunately, an a priori determination of a "trust" region beyond which the approximate surface poorly represents the true surface requires a knowledge of higher derivatives (e.g., of third and higher order for the quadratic approximation used here). An inappropriate ad hoc determination of such a "trust" region could result in a significant increase of computational effort or possible divergence.

We describe here a dynamic step reduction procedure to deal with the overstepping problem. This procedure is an extension of Fletcher's algorithms and is employed in the numerical examples treated in the next section. For a given step $\mathbf{x}$, the size of the relative error incurred by the model function $E(\mathbf{x})$ over the true function $\tilde{E}(x)$ is $\delta(\mathbf{x}) = |\Delta\tilde{E} - \Delta E|/|\Delta E|$ where $\Delta\tilde{E} = \tilde{E}(x) - \tilde{E}_0$ is the difference in the true function for the step given by $\mathbf{x}$. A "trust region" can be defined by those steps $\mathbf{x}$ for which the relative error is less than a specified amount $\delta(\mathbf{x}) \leq \tilde{\delta}$, where $\tilde{\delta}$ is a predetermined constant fraction. A step $\mathbf{x}$ that fails to satisfy this criterion is judged to be too long and its length must be reduced.

The requisite reduction in step length can be obtained by performing a one-dimensional line search along $\mathbf{x}_1 = \mathbf{x}_0 + t\boldsymbol{u}$, where $\mathbf{u} = \mathbf{x}/|\mathbf{x}|$ is a unit vector lying along the step $\mathbf{x}$ and the parameter t $(0 \leq t \leq |\mathbf{x}|)$ defines the line. The *scalar* values of the gradient and Hessian along this line at $\mathbf{x}_0$ are $g = \mathbf{g}^+\mathbf{u}$, $h = \mathbf{u}^+\boldsymbol{H}\mathbf{u}$. Then the values of t which satisfy $\delta(t) \leq \tilde{\delta}$ can be shown to be those that obey

$$Mt^2/6 \leq \tilde{\delta}|g + \tfrac{1}{2}ht| \tag{17}$$

where M is (an estimate of) the largest absolute value which the third derivative of the $\tilde{E}$ function assumes along the line $0 \leq t \leq |\mathbf{x}|$. We can estimate M using a Taylor expansion about $t = 0$: $M \cong 6|\tilde{\Delta}E - \Delta E|/|\mathbf{x}|^3$. The solution of eq 17 can be straightforwardly obtained for particular values of g, h, and M calculated at the point $\mathbf{x}_0$. The application of this procedure is straightforward since it is performed on a one-dimensional curve which requires only evaluation of the value of the function $\tilde{E}(x)$ at the new point $\mathbf{x}$. We have made use of this step-reduction algorithm for *all* of the numerical applications reported in this paper. The choice for the $\tilde{\delta}$ parameter $\tilde{\delta} = 0.3$ was found, after a little experimentation, to provide reasonable and seemingly reliable line searches when employed in this algorithm.

Using Approximate Hessians. So far we have developed and analyzed the expected performance of two RFO schemes assuming that the elements of the gradient $\mathbf{g}$ and Hessian $\boldsymbol{H}$ are available. A very important application of these methods arises in potential energy surfaces of molecules. In such cases, the ab initio calculation of the Hessian matrix elements are tedious and laborius, primarily because of the large number of integral derivatives which contribute to them. It is therefore important to explore the performance of the above RFO schemes in which approximate (and presumably easier to evaluate) Hessian matrix elements are used. The methods proposed in the literature for the evaluation of approximate or updated Hessians are based upon finite-difference approximations based upon the following equation

$$\boldsymbol{H}_{K+1}(\mathbf{x}_{K+1} - \mathbf{x}_K) = (\mathbf{g}_{K+1} - g_K) \tag{18}$$

involving gradient elements from two successive steps $\mathbf{x}_K$ and $\mathbf{x}_{K+1}$.

We have used the so called Powell update procedure[11] because it preserves the symmetric character of $\boldsymbol{H}_{K+1}$, although it does not force $\boldsymbol{H}_{K+1}$ to have the same eigenvalue structure as $\boldsymbol{H}_K$. This means that the Hessian is free to evolve an eigenvalue structure which changes during the energy surface walk, which is clearly an essential feature of any acceptable update method. The Powell update expression is given as follows

$$\boldsymbol{H}_{K+1} = \boldsymbol{H}_K + (\mathbf{S}^+\mathbf{S})^{-1}\left\{\mathbf{V}\mathbf{S}^+ + \mathbf{S}\mathbf{V}^+ - \frac{(\mathbf{V}^+\mathbf{S})(\mathbf{S}\mathbf{S}^+)}{(\mathbf{S}^+\mathbf{S})}\right\} \tag{19}$$

where

$$\mathbf{S} = \mathbf{x}_{K+1} - \mathbf{x}_K, \quad \text{and} \quad \mathbf{v} = (\mathbf{g}_{K+1} - \mathbf{g}_K) - \boldsymbol{H}_K\mathbf{S} \tag{20}$$

This process of recursive updating requires an initial Hessian $\boldsymbol{H}_0$. We have formed $\boldsymbol{H}_0$ by carrying out a finite-difference gradient calculation on a grid of points near the starting point $\mathbf{X}_0$. The applications presented here demonstrate that one can locate desired stationary points using RFO schemes with updated Hessian. We find that, due to the approximate nature of the Hessian, a step restrictive procedure (such as one presented earlier) should be used along with the updated Hessian to mainly avoid an occasional large misleading step.

Applications

We have tested the RFO algorithms on model surfaces. We have used the step modification algorithm, described earlier, throughout in search of the various stationary points on surfaces presented here. The first application involves searches for both a minimum and a saddle point on the following model potential surface:

$$E(x_1,x_2) = (a - bx_2)x_1^2e^{-x_1^2} + (c/2)x_2^2 \tag{21}$$

with $a = b = c = 1$. Since the surface (see Figure 3) has relatively similar shapes in both directions, we chose to scale the x and y axes both by the same amount in the dimensionless units of this model surface. This same surface was examined earlier by Cerjan and Miller[4] (CM) and by Simons et al.[3] Its minimum is located at (0,0) and two identical saddle points occur at (±1,0). For minimum searches, both algorithms reduce to the RFO case where we expect the algorithm to tend to minimize along each principal mode. Figure 3 shows the locus of the RFO steps starting from near the saddle point and following the "valley" along the x axis to the minimum. The RFO path follows the valley hugging the valley floor. Figure 3 also shows the "walks" of the RFO and P-RFO algorithms for the saddle-point search starting near the minimum utilizing both analytical and updated Hessians. The analytical-Hessian P-RFO walk hugs the valley floor more closely than does the corresponding RFO walk. Starting from near the minimum at (0,0), the soft mode (corresponding to the lowest eigenvalue of the Hessian) produces steps along the y axis. At the fourth step, in the region near (0.01,0.7), both eigenvalues of the Hessian are nearly equal and are positive. This situation induces a P-RFO step with nearly equal x and y components, due to the independent (partitioned) calculation of each component. As a result, this P-RFO step cuts quite perpendicular to the potential contour lines as a valley-walk should do. In contrast, the RFO step gives a larger component along the x direction which

(11) Powell, M. J. D. *Math. Prog.* **1971**, *1*, 26.

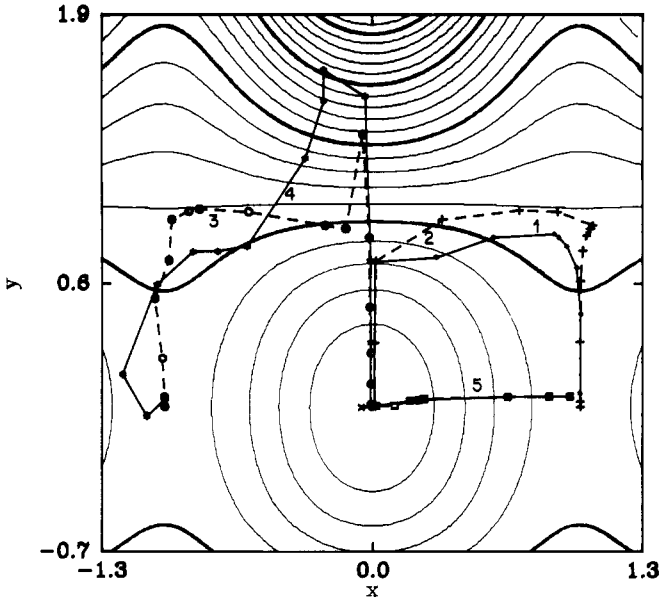

Figure 3. Walks on the CM potential surface: (1–4) correspond to walks toward a saddle point (0,±1) starting near the minimum at (0,0); (1) RFO; (2) P-RFO; (3) RFO with updated Hessian; (4) P-RFO with updated Hessian; (5) RFO walk to the minimum.

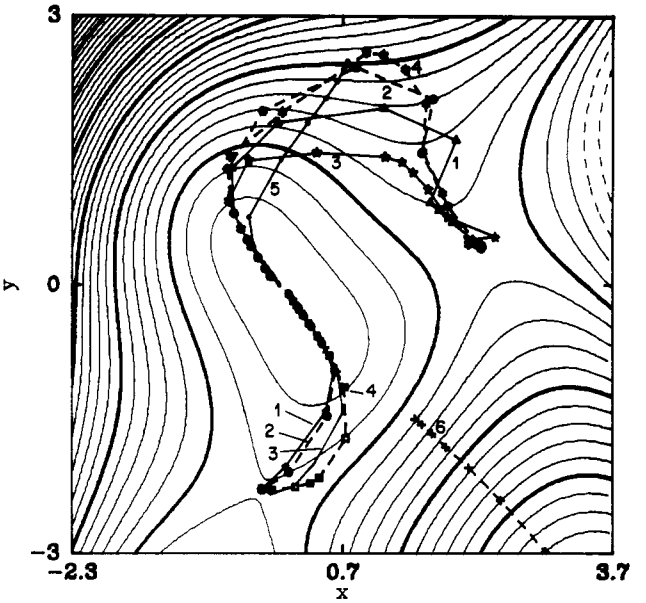

Figure 4. Walks on the Adams potential surface: (1–4) correspond to walks toward saddle points starting near the minimum at (0,0); (1) RFO; (2) P-RFO; (3) RFO with updated Hessian; (4) P-RFO with updated Hessian; (5) RFO walk to the minimum, (6) RFO walk toward the maximum.

results in a step less perpendicular to the contours and hence further from the valley floor. For the next four steps, the soft mode lies very much along the x direction, and both the RFO and P-RFO methods perform rather well. In the region near (1.0,0.8) the soft mode makes a nearly right angle turn as do the steps of both the RFO and P-RFO methods.

The updated-Hessian paths start out following the soft mode up the y axis but do not take the turn to proceed along the x axis in the region near (0.01,0.7). This is due to the insufficiently accurate (updated) Hessian's inability to describe the change in the nature of the local soft mode. However, after a few additional steps, the updated-Hessian walk has gained enough data to turn around completely and follow the appropriate (x) valley to the saddle point.

The second surface used for testing the RFO walks is the Adams potential[12]

$$V(x,y) = 2x^2(4 - x) + y^2(4 + y) - xy(6 - 17e^{-r^2/4})$$

which has a minimum at (0,0), a maximum at (3.8239,−4.4096), and two saddle points of first order at (2.4104,0.4419) and (−0.1985,−2.2793). This surface is also dimensionless so the choice $a = 1$ is again appropriate. Figure 4 shows all of our walks to various stationary points; the four walks to each of the two saddle points (starting near the minimum) correspond to the paths of the RFO and P-RFO algorithms using analytical and updated Hessians. One walk each to the maximum and to the minimum are also depicted in which the analytical Hessian was employed. The updated-Hessian walks for these latter two cases (not shown) are almost indistinguishable from the analytical-Hessian walks. In viewing the various saddle-point walks, we note that the valleys leading to each saddle are curved channels; this is very much in contrast to the CM surface. All of the walks to the saddle points start out following the soft-mode valley at an angle of approximately 45°, then turn along with the soft-mode valley to the appropriate saddle point. The P-RFO walks again follow the valley bottoms closer than do the corresponding RFO walks, as expected. Unlike their performance on the CM surface, the updated-Hessian walks follow paths very close to the analytical Hessian paths.

(12) Adams, N., unpublished results.

Conclusions

From the above comparative analysis of the two RFO algorithms presented here, we believe the P-RFO algorithm to be most efficient. It performs extremizations with respect to two sets of variables and calculates separately two independent λ_n and λ_p parameters for the minimization and maximization modes, respectively. The original RFO algorithm is among the simplest to apply and, unlike its P-RFO variant, it does not require a basis transformation to enable partitioning of the variables for maximization and minimization. Both of the RFO algorithms are also well suited for locating stationary points of *large* dimensional surfaces (for example, for larger biologically important molecules) since only the low eigensolutions of the RFO matrices need be calculated. We have previously used a method similar to the P-RFO algorithm for the optimization of MCSCF wave functions[7,8] for which one always needs to minimize the energy in the orbital variable (for any electronic state) as a result of which there exists a natural partitioning into two classes of variables (orbital and configuration).

Acknowledgment. We acknowledge many stimulating discussions with Jim Jensen and Hugh Taylor and Judy Ozment for her help for drawing contour graphs. We also acknowledge the financial support of the National Science Foundation (Grant 8206845) and the donors of the Petroleum Research Fund, administered by the American Chemical Society. Partial support for R.S. was provided by the Office of Basic Energy Sciences, Division of Chemical Sciences, U.S. Department of Energy, under Contract W-31-109-ENG-38.