

## Following gradient extremal paths\*

**H. Bernhard Schlegel**

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

Received June 3, 1991/Accepted September 17, 1991

**Summary.** For any point on a gradient extremal path, the gradient is an eigenvector of the hessian. Two new methods for following the gradient extremal path are presented. The first greatly reduces the number of second derivative calculations needed by using a modified updating scheme for the hessian. The second method follows the gradient extremal using only the gradient, avoiding the hessian evaluation entirely. The latter algorithm makes it possible to use gradient extremals to explore energy surfaces at higher levels of theory for which analytical Hessians are not available.

**Key words:** Gradient extremals – Reaction paths – Saddle points – Potential energy surfaces

### 1 Introduction

The investigation of energetics and reactivity by quantum chemical methods involves the exploration of potential energy surfaces. The features of interest of these surfaces include minima, saddle points and the paths that connect them. The minima, representing reactants, products and intermediates, are relatively easy to find by gradient based optimization methods (for a review of geometry optimization methods, see [1]). A transition state is more difficult to find [1], since it is first order saddle point (i.e., a maximum in one direction and a minimum in all other directions). Reaction paths can be found by following the steepest descent path from the transition states to the reactants and products [2]. If the steepest descent or minimum energy path (MEP) is computed in mass weighted cartesian coordinates, it is termed the intrinsic reaction coordinate (IRC) [2]. A number of algorithms are available for following reaction paths [3–7]. Generally, these methods are unsuitable for following the path uphill from the reactants to the transition state. In part, this is because there are no local criteria for determining whether or not a point is on the minimum energy path connecting the transition state and the minimum. An alternative to the minimum

---

\* Dedicated to Prof. Klaus Ruedenberg

energy path is the gradient extremal path [8–13], i.e., the path along which the gradient is an eigenvector of the hessian:

$$\dot{\tilde{x}}(s) \text{ such that } \mathbf{H}(s)\dot{\tilde{x}}(s) = \lambda(s)\dot{\tilde{x}}(s) \quad (1)$$

where  $s$  is the arc length along the path,  $\dot{\tilde{x}}(s)$  is the gradient at  $\tilde{x}(s)$  and  $\mathbf{H}(s)$  is the hessian at  $\tilde{x}(s)$ . The gradient extremal path is locally defined – from any point on the energy surface one can step to the gradient extremal path. Algorithms have been devised for following gradient extremal paths (see [11] and references cited), and these methods work equally well in the uphill and the downhill directions. However, gradient extremal paths have some drawbacks: (a) they have no dynamic or mechanistic significance, (b) they do not always follow the most direct route from a minimum to a transition state (see [10] for an example) and (c) existing algorithms require second derivatives at each step. Second derivative calculations are significantly more expensive than gradient computations. Even more problematic is the fact that analytic second derivatives are not available for most of post-SCF levels of theory that are currently being used in accurate studies of potential energy surfaces. The present paper outlines two algorithms for following gradient extremals: one method that requires fewer second derivative calculations and a second method that requires no second derivative evaluations.

## 2 Theory

It has been shown that the gradient is not tangent to the gradient extremal path [10]. This can be seen readily by writing  $\tilde{x}(s)$  as a Taylor expansion and by differentiating Eq. (1) with respect to the arc length,  $s$ .

$$\tilde{x}(s) = \tilde{x}(0) + s\dot{\tilde{x}}^0(s) + \frac{1}{2}s^2\ddot{\tilde{x}}^1(s) + \dots; \quad \dot{\tilde{x}}^0(s) = \frac{\partial \tilde{x}}{\partial s}; \quad \ddot{\tilde{x}}^1(s) = \frac{\partial^2 \tilde{x}}{\partial s^2} \quad (2)$$

$$\frac{\partial \mathbf{H}}{\partial \tilde{x}} \frac{\partial \tilde{x}}{\partial s} \dot{\tilde{x}} + \mathbf{H} \frac{\partial \dot{\tilde{x}}}{\partial \tilde{x}} \frac{\partial \tilde{x}}{\partial s} = \frac{\partial \lambda}{\partial s} \dot{\tilde{x}} + \lambda \frac{\partial \dot{\tilde{x}}}{\partial \tilde{x}} \frac{\partial \tilde{x}}{\partial s} \quad (3)$$

$$(\mathbf{F}\dot{\tilde{x}}^0)\dot{\tilde{x}} + \mathbf{H}^2\dot{\tilde{x}}^0 = \frac{\partial \lambda}{\partial s} \dot{\tilde{x}} + \lambda \mathbf{H}\dot{\tilde{x}}^0 \quad (4)$$

where  $\partial \mathbf{H} / \partial \tilde{x} = \mathbf{F}$  is the third derivative,  $\partial \dot{\tilde{x}} / \partial \tilde{x} = \mathbf{H}$  is the hessian and  $\partial \tilde{x} / \partial s = \dot{\tilde{x}}^0$  is the tangent to the gradient extremal path. The tangent can be found by solving the set of linear equations given in Eq. (4). If the third derivative term is non-zero, the tangent cannot be an eigenvector of the hessian. Equation (1) requires the gradient to be an eigenvector of the hessian; hence, the tangent differs from the gradient by a term depending on the third derivatives. Thus, any attempt to follow the gradient extremal path by simply stepping along the gradient is incorrect even in the first order term in Eq. (2).

The gradient extremal following algorithm of by Jørgensen, Jensen, and Helgaker [11] does yield the correct tangent (in the modification given below), since an optimization step is added to return to the gradient extremal path after stepping along the gradient:

$$\begin{aligned} \tilde{x}(s') &= \tilde{x}(0) + \sigma \dot{\tilde{x}}(0) \\ \tilde{x}(s) &= \tilde{x}(s') - (\mathbf{I} - \dot{\tilde{x}}(s) \cdot \dot{\tilde{x}}(s)^T) \mathbf{H}(s)^{-1} \dot{\tilde{x}}(s) \end{aligned} \quad (5)$$

where  $\tilde{g}(0)$  satisfies  $\mathbf{H}(0)\tilde{g}(0) = \lambda(0)\tilde{g}(0)$  and  $\tilde{v}(s)$  is the appropriate eigenvector of  $\mathbf{H}(s)$ .

The use of Eq. (5) for following gradient extremal paths can be quite costly, since the hessian is required at each step. For SCF methods, the analytical computation of the hessian is usually several times more expensive than the gradient calculation. For many post-SCF methods, the hessian must be computed by numerically differentiating the gradient. This can become prohibitively expensive as the number of atoms/degrees of freedom increases. One alternative is to calculate the hessian at only a few points along the gradient extremal path and to use a hessian updating formula for the intermediate points. The appropriate formulas can be found by expanding the gradient and hessian in a Taylor series:

$$g_i^1 = g_i^0 + \sum_j \mathbf{H}_{ij}^0 \Delta x_j + \frac{1}{2} \sum_{j,k} \mathbf{F}_{ijk}^0 \Delta x_j \Delta x_k + \dots \quad (6)$$

$$\mathbf{H}_{ij}^1 = \mathbf{H}_{ij}^0 + \sum_k \mathbf{F}_{ijk}^0 \Delta x_k + \dots \quad (7)$$

The DFP updating formula (see [14] or any other text on unconstrained optimization methods) for the hessian is:

$$\hat{\mathbf{H}}^1 = \mathbf{H}^0 + \Delta \mathbf{H}; \quad \Delta \mathbf{H} = \frac{\Delta \tilde{g} \Delta \tilde{g}^t}{\Delta \tilde{g}^t \Delta \tilde{x}} - \frac{\mathbf{H} \Delta \tilde{x} \Delta \tilde{x}^t \mathbf{H}}{\Delta \tilde{x}^t \mathbf{H} \Delta \tilde{x}}; \quad \Delta \tilde{g} = \tilde{g}^1 - \tilde{g}^0 \quad (8)$$

This and related updating formulas, such as BFGS, OC, etc. [14] all have the property:

$$\Delta g_i = \sum_j \hat{\mathbf{H}}_{ij}^1 \Delta x_j = \sum_j \mathbf{H}_{ij}^0 \Delta x_j + \frac{1}{2} \sum_{j,k} \mathbf{F}_{ijk}^0 \Delta x_j \Delta x_k \quad (9)$$

or

$$\Delta \tilde{g} = \hat{\mathbf{H}}^1 \Delta \tilde{x} = \mathbf{H}^0 \Delta \tilde{x} + \Delta \mathbf{H} \Delta \tilde{x}$$

However, the correct expansion of  $\mathbf{H}^1$  needed for Eq. (5) has the following component along  $\Delta \tilde{x}$ :

$$\sum_j \mathbf{H}_{ij}^1 \Delta x_j = \sum_j \mathbf{H}_{ij}^0 \Delta x_j + \sum_{j,k} \mathbf{F}_{ijk}^0 \Delta x_j \Delta x_k \quad (10)$$

By comparison with Eq. (9), we can see that the third derivative term in Eq. (10) can be approximated by  $2 \Delta \mathbf{H} \Delta \tilde{x}$ . This leads to the following update formula for the hessian at  $\tilde{x}^1$  given the exact hessian  $\mathbf{H}^0$  at  $\tilde{x}^0$ :

$$\mathbf{H}^1 = \mathbf{H}^0 + 2 \Delta \mathbf{H} \quad (11)$$

The extra factor of 2 arises from the fact that Eq. (8) is required to fit the change in the gradient with an average hessian, whereas Eq. (11) estimates the change in the hessian. This updating formula can be used with the Jørgensen, Jensen, and Helgaker algorithm [11] to following gradient extremal paths with a greatly reduced number of hessian calculations. Because Eq. (11) updates the hessian using only the information from stepping in the direction of  $\Delta \tilde{x}$ , it is probably advisable to re-compute the hessian every 5 or 10 steps.

An alternate algorithm for following gradient extremal paths can be devised that relies only on the gradient and avoids the direct calculation of the hessian. Consider  $\tilde{x}^1$  on the gradient extremal path and  $\tilde{x}^*$  a small displacement  $\tau$  from  $\tilde{x}^1$  along the gradient  $\tilde{g}^1$ .

$$\tilde{x}^* = \tilde{x}^1 + \tau \tilde{g}^1 / |\tilde{g}^1| \quad (12)$$

$$\begin{aligned} \tilde{g}^* &= \tilde{g}^1 + \mathbf{H}^1(\tilde{x}^* - \tilde{x}^1) + O(\tau^2) = \tilde{g}^1 + (\tau/|\tilde{g}^1|)\mathbf{H}^1\tilde{g}^1 + O(\tau^2) \\ &= \tilde{g}^1 + (\tau/|\tilde{g}^1|)\lambda\tilde{g}^1 + O(\tau^2) = (1 + \lambda\tau/|\tilde{g}^1|)\tilde{g}^1 + O(\tau^2) \end{aligned} \quad (13)$$

where the condition for being on the gradient extremal,  $\mathbf{H}^1\tilde{g}^1 = \lambda\tilde{g}^1$ , has been used to simplify the last equation. Thus, if  $\tilde{x}^1$  is on the gradient extremal,  $\tilde{g}^*$  is parallel to  $\tilde{g}^1$ . This can be rewritten as:

$$\tilde{g}^* - (\tilde{g}^{*\prime}\tilde{g}^1)\tilde{g}^1/|\tilde{g}^1|^2 = \vec{0} \quad (14)$$

to first order in  $\tau$ . Note that Eq. (14) is a local condition for the gradient extremal path that is independent of the hessian and requires only two gradient evaluations. The point  $\tilde{x}^1$  can be found by a constrained optimization such that Eq. (14) is satisfied and the stepsize from a previous point on the path is constant, i.e.:

$$|\tilde{x}^1 - \tilde{x}^0| = \sigma \quad (15)$$

This is analogous to the Müller–Brown [4] and Gonzalez–Schlegel [7] reaction path following algorithms with the residual from Eq. (14) used in place of the gradient in the constrained optimization. Since Eq. (14) is a local criterion ( $\tau \ll \sigma$ ),  $\tilde{x}^1$  will be on the gradient extremal path, regardless of the stepsize  $\sigma$ ; hence, the question of the formal order of the algorithm is not relevant. Because two gradients must be calculated to evaluate Eq. (14), following gradient extremal paths will be roughly twice as expensive as following minimum energy paths. The payoff, of course, is that gradient extremal paths can be followed uphill to the transition state, whereas minimum energy paths, in general, cannot. One drawback of the gradient method for following extremal paths is that bifurcations cannot be detected readily.

Minimum energy paths have the desirable feature of being the most direct route between the transition state and reactants and products (in the sense of greatest energy lowering for distance traveled). Gradient extremal paths tend to be somewhat less direct (see [10] for an example). Some insight into this can be gained by comparing the two paths. For a minimum energy path, the tangent and curvature for the Taylor expansion of the path given in Eq. (2) are:

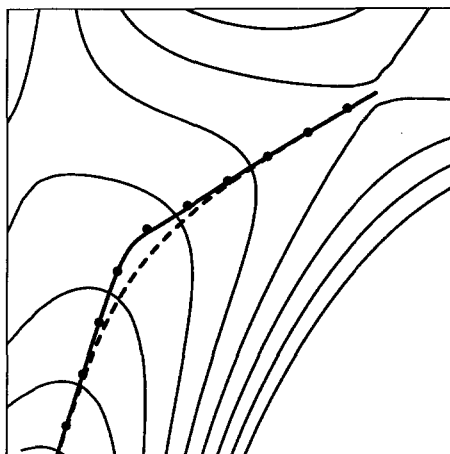
$$\tilde{v}^0 = -\tilde{g}/|\tilde{g}|; \quad \tilde{v}^1 = (\mathbf{H}\tilde{v}^0 - (\tilde{v}^{0\prime}\mathbf{H}\tilde{v}^0)\tilde{v}^0)/|\tilde{g}| \quad (16)$$

The initial curvature for a minimum energy path starting from any point on the gradient extremal path is zero. Since  $\tilde{g}$  on the gradient extremal path is an eigenvector of the hessian, then so is  $\tilde{v}^0$ ; substitution of the expression for  $\tilde{v}^0$  into  $\tilde{v}^1$  simplifies to zero. This suggests, but does not prove, that the largest deviation between a minimum energy path and the corresponding gradient extremal path will occur near regions of large curvature in the minimum energy path.

### 3 Application

A simple application of the gradient extremal path following algorithm given by Eqs. (14) and (15) is illustrated in Fig. 1. This is a portion of a model surface proposed by Ruedenberg et al. [10]:

$$E(x, y) = \frac{1}{2}(xy^2 - yx^2 + x^2 + 2y - 3) \quad (17)$$



**Fig. 1.** Gradient extremal and minimum energy paths on the surface given by Eq. (17) ( $-1 \leq x \leq 3$ ,  $-2 \leq y \leq 2$ ). *Dashed line* – minimum energy path, *solid line* – gradient extremal path, *dots* – points on the gradient extremal path calculated using Eqs. (14) and (15)

A complete analysis of the various gradient extremal and minimum energy paths can be found in [10]. The dashed line, the minimum energy path, was calculated by steepest descent from the transition state using Euler's method (stepsize = 0.01). The gradient extremal path was followed uphill using the algorithm of Jørgensen, Jensen, and Helgaker [11] (Eq. (5), stepsize = 0.02). The dots represent the points found by applying Eqs. (14) and (15) to follow the gradient extremal path uphill (stepsize  $\sigma = 0.5$ ,  $\tau = 0.005$ ). All calculations for this model surface were carried out with Mathematica [15]. The method proposed in this paper follows the gradient extremal path quite well for this simple example. The constrained minimization of Eq. (14) is rather easy in two dimensions. For higher dimensions, techniques similar to the reaction path following methods [7] will be needed.

*Acknowledgements.* This work was supported by NSF grant CHE 90-20398. The author would like to thank a referee of this grant as well as the organizers of the symposium in honor of Professor Ruedenberg for drawing my attention to gradient extremals.

## References

- Schlegel HB (1987) *Adv Chem Phys* 67:249
- Fukui K (1981) *Acc Chem Res* 14:363
- Ishida K, Morokuma K, Komornicki A (1977) *J Chem Phys* 66:2153
- Müller K, Brown LD (1979) *Theor Chim Acta* 53:75
- a. Page M, McIver JW (1988) *J Chem Phys* 88:922  
b. Page M, Doubleday C, McIver JW (1990) *J Chem Phys* 93:5634
- a. Garrett BC, Redmon MJ, Steckler R, Truhlar DG, Baldrige KK, Bartol D, Schmidt MW, Gordon MS (1988) *J Phys Chem* 92:1476  
b. Koseki S, Gordon MS (1988) *J Phys Chem* 93:118
- a. Gonzalez C, Schlegel HB (1989) *J Chem Phys* 90:2154  
a. Gonzalez C, Schlegel HB (1990) *J Phys Chem* 94:5523  
a. Gonzalez C, Schlegel HB (1991) *J Chem Phys* 95:5853
- Panicir J (1975) *Collect Czech Chem Commun* 40:1112
- Basilevsky MV, Shamov AG (1981) *Chem Phys* 60:347

10. Hoffman DK, Nord RS, Ruedenberg K (1986) *Theor Chim Acta* 69:265
11. Jørgensen P, Aa Jensen HJ, Helgaker T (1988) *Theor Chim Acta* 73:55
12. Quapp W (1989) *Theor Chim Acta* 75:447
13. Shida N, Almlöf JE, Barbara PF (1989) *Theor Chim Acta* 76:7
14. Scales LE (1985) *Introduction to non-linear optimization*. Macmillan, New York
15. Wolfram S (1988) *Mathematica*. Addison-Wesley, New York