Computers Chem. Vol. 17, No. 1, pp. 21–25, 1993
Printed in Great Britain.

0097-8485/93 $6.00 + 0.00

#4

PROCEDURE FOR DETERMINING DYNAMICALLY DEFINED REACTION PATH

L. L. STACHÓ[1] and M. I. BÁN[2]*
[1]JATE Bolyai Institute for Mathematics, Aradi Vértanúk tere 1, H-6720 Szeged, Hungary and
[2]JATE Institute of Physical Chemistry, P.O. Box 105, H-6701 Szeged, Hungary

(Received 7 October 1992; in revised form 6 December 1992)

Abstract—A practical computational program and pertinent sections of the code illustrated by flow-charts for a searching procedure to determine dynamically defined reaction paths have been discussed.

1. INTRODUCTION

Considerable attention has recently been paid to improve methods of finding reaction paths and location of stationary points corresponding to equilibrium and transition states on potential energy hypersurfaces. In relation to our theoretical work (Stachó & Bán, 1992a) from among the hundreds of references on the same subject we would like to highlight those of Silver (1966), Fukui (1970), McIver & Komornicki (1974), McCullogh & Silver (1975), Fukui *et al.* (1975), Halgren *et al.* (1975), Ishida *et al.* (1977), Halgren & Lipscomb (1977), Peterson & Csizmadia (1977), Mezey *et al.* (1977), Tachibana & Fukui (1978, 1979), Muller & Brown (1979), Lawley (1980), Scharfenberg (1980), Truhlar (1981), Fukui (1981), Sana *et al.* (1981), Bell & Crighton (1984), Quapp & Heidrich (1984), Murrell *et al.* (1984), Jorgensen & Simmons (1985), Hirst (1985), Baker (1986, 1987), Schlegel (1987), Mezey (1987, 1988), Baker & Gill (1988), Jasien & Shepard (1988), Kliesch *et al.* (1988), Cummins & Gready (1989), Mezey (1990), Kraka & Dunning (1990), Xantheas *et al.* (1991), Ischtwan & Collins (1991), Natanson *et al.* (1991), Chielke & Truhlar (1991), Gonzalez & Schlegel (1991), Billing (1992), Melissas & Truhlar (1992), and Datta (1992).

In the first paper of our series the fundamental theory (Stachó & Bán, 1992a), and in the second paper (Stachó & Bán, 1992b) the mathematical background of a global strategy for determining dynamically defined reaction path (DDRP) has been given. In our third paper (Stachó & Bán, 1993) the algorithm and aspects of numerical applications have been described. In the present paper the characteristic features of a practical computational program based on the theory laid in the first three papers of this series and pertinent sections of the code illustrated by flow-charts have been discussed. A next paper (Dömötör *et al.*, 1993b) will show the use of the searching procedure of DDRP on examples of simple chemical reactions, by employing quantum chemical methods for calculating the (energy) functional. Here, also experiences gained by the implementations of quantum chemical procedures into the program of DDRP will be discussed.

*Author for correspondence.

2. PROCEDURE

Recalling the fundamental theorem and the theoretical considerations described in the first paper (Stachó & Bán, 1992a) we suggest to determine tangent curves of a vector field V between two given singular points by taking the limit (in the sense of Hausdorff distance) for $t \to \infty$ of curves of the form

$$c^t := \exp(tV)(c^0). \tag{1}$$

Here c^0 is any piecewise analytic curve joining the given singularities or simply joining points from the respective catchment regions of these singularities. The exponentials of the vector field V are the mappings $x \mapsto x^t$ satisfying the steepest descent path equation

$$\frac{\mathrm{d}}{\mathrm{d}t} x^t = V(x^t), \quad x^0 = x \quad \exp(tV)(x) := x^t.$$

Thus in this manner we approach the reaction path between two singularities of the potential energy function by not trying to move forward in the suspected reaction valley according to some fixed strategy (as most known popular methods do), but—using a well-visualizable analogy—by "stretching a cord onto the reaction path" from an almost arbitrary starting position joining the given two endpoints. A careful mathematical analysis (Stachó & Bán, 1992b) ensures the convergence of this procedure even in cases with a seemingly very hopeless starting curve or a potential surface with the possibility of a ramifying system of intermediate products. Of course, this high grade of reliability must be payed through a higher amount of calculations. However, as it was pointed out (Stachó & Bán, 1993) our

method is completely parallelizable, in contrast with other strategies of moving forward. Moreover it is very robust in the sense that it is not very sensible for the local accuracy of the chosen method of solving the differential equation $(\mathrm{d}/\mathrm{d}t)x^t = V(x^t)$. Indeed, the simplest method of steps (with relatively large steps) is suitable for our practical purposes.

The main features of our algorithm are as follows. First we fix a curve c^0 (say a straight line segment) between the points representing reactants and products. It is not necessary to know very exactly where these points are because the algorithm will find them automatically when the endpoints of a curve lie in their respective catchment regions. We fix also a number T which will play the role of a (virtual) time unit. In the first step we determine the curve c^T by solving the differential equation $(\mathrm{d}/\mathrm{d}t)x^t = V(x^t)$ on the time interval $[0, T]$ with all possible initial values $x^0 \in c^0$. Knowing the curve c^T, we determine c^{2T} in a similar manner and so we continue. For $n \to \infty$ the curves c^{nT} converge uniformly to an IRC, i.e. to a curve consisting of steepest descent path pieces between stationary points and joining continuously the points of reactants and products.

Next we describe the numerical realization of the above somewhat idealized procedure in details.

As usual, we represent curves by finite sequences of their points. The main novelty of our algorithm is the application of a homogenization $\mathscr{H}$ by the aid of which we construct an approximate representative $Z = \{z_1, \dots, z_m\}$ for the curve $c^{t+\delta}$ from an other approximate representative $Y = \{y_1, \dots, y_l\}$ of c^t in the form

$$Z = \mathscr{H}(\Phi^\delta(Y)) \tag{2}$$

where Φ^δ denotes some numerical approximation of $\exp(\delta V)$. It turns out that even the relatively simplest constructions for $\mathscr{H}$ and Φ^δ may lead to satisfactory solutions. For Φ^δ we can use the ordinary step approximations of $\exp(\delta V)$. That is, with a given input parameter M, we compute

$$\Phi^\delta(y) = \Phi^\delta_M = \underbrace{F_{\delta/M}(F_{\delta/M}(\cdots(F_{\delta/M}(y)\cdots)))}_{M \text{ times}} \tag{3}$$

where

$$F_h(y) := y + hV(y) \quad (h > 0, y \in \mathbb{R}^n). \tag{4}$$

It may be convenient to restriet the resulting curves to some sphere $D := \{y \in \mathbb{R}^n : \|y\| \leqslant R\}$ where the radius R is again an input parameter. Then we can use the transformations

$$\tilde{F}_h := P(F_h) \tag{5}$$

instead of F_h with

$$P := [\text{projection onto } D]. \tag{6}$$

With an appropriately chosen parameter ϵ we can use the following homogenization procedure: $\mathscr{H} = \mathscr{H}_\epsilon$. To calculate

$$\{z_1, \dots, z_m\} = \mathscr{H}_\epsilon\{y_1, \dots, y_l\} \tag{7}$$

first we determine the distances

$$d_i := \|y_{i+1} - y_i\| \quad (i = 1, \dots, l-1). \tag{8}$$

Then we choose consecutively the indices

$$1 = i_1 < i_2 < \cdots < i_p = n \tag{9}$$

defined unambiguously by the requirement

$$\sum_{i_k \leqslant i < i_{k+1}-1} d_i \leqslant \epsilon < \sum_{i_k \leqslant i < i_{k+1}} d_i \quad (1 \leqslant k < p-1). \tag{10}$$

Finally we obtain the sequence $\{z_1, \dots, z_m\}$ from $\{y_{i_k} : k = 1, \dots, p\}$ by deleting all points with indices not belonging to the set $\{i_1, \dots, i_p\}$ and by inserting entier (d_{i_k}/ϵ) affinely interpolating equidistant points between y_{i_k} and $y_{i_{k+1}}$ whenever $i_{k+1} = i_k + 1$ (i.e. if $d_{i_k} > \epsilon$).

Thus, given the controlling parameters δ, M, ϵ (and possibly the radius R of cutting) with a representative sequence $Y^0 = \{y^0_i := 1, \dots, l_0\}$ of the initial curve c^0, we compute iteratively the representative sequences

$$Y^j = \mathscr{H}_\epsilon(\Phi^\delta_M(Y^{j-1})) \tag{11}$$

for the curves $c^{j\delta}$ $(j = 1, 2, \dots)$.

A reasonable stop-condition can be given in terms of the norm maximum of the vector field V. Introducing a new controlling parameter λ, a satisfactory stop-condition for the idealized version of the algorithm is the condition

$$d_H(c^{j\delta}, c^{(j-1)\delta}) < \lambda\delta \max_x \|V(x)\| \tag{12}$$

where d_H stands for Hausdorff distance of subsets in $\mathbb{R}^n$. However, no *a priori* upper estimates for $\|V\|$ are available in the literature. On the other hand, even if $\max_x \|V(x)\|$ is known, the simplest analogous condition

$$d_H(\overline{Y^j}, \overline{Y^{j-1}}) < \lambda\delta \max_x \|V(x)\| \tag{13}$$

cannot be satisfied in most cases where $\overline{Y}$ denotes the polygon joining the consecutive points of Y. The following may indeed happen: all the points of Y^j and Y^{j-1} lie very near $c^\infty = \lim_{t\to\infty} c^t$ and a singularity y of V at a sharp angle of c^∞ belongs to Y^{j-1} while the homogenization $\mathscr{H}_\epsilon$ deletes y from Y^j. In this case $d_H(\overline{Y^j}, \overline{Y^{j-1}})$ may exceed the value $\epsilon/4$.

A way of avoiding this difficulty is to modify somewhat the homogenization $\mathscr{H}_\epsilon$. Let us change the definition of the distances d_i as follows. Instead of setting $d_i = \|y_{i+1} - y_i\|$ define

$$d_i = \|y_{i+1} - y_i\| + \frac{\epsilon}{\lambda}\,\mathrm{angle}(y_i - y_{i-1}, y_{i+1} - y_i)$$
$$(1 < i < l). \tag{14}$$

In this manner any stop-condition of the form

$$d_H(Y^j, Y^{j-1}) < \Omega \tag{15}$$

can be reached when choosing λ sufficiently small and M sufficiently large (under the hypothesis of the theorem (Stachó & Bán, 1993)).

Small values of λ or ϵ may lead to representatives Y^j consisting of too many points while choosing large values of M causes time-consuming calculations with relatively little gain in accuracy.

For practice we suggest that the above described algorithm should be divided into two procedures (see the flow-charts in Figs 1 and 2):

NEWCURVE calculates the curve representative $Z = \Phi^\delta_m(Y)$ for $c^{t+\delta}$ from a representative Y of c^t in which a subroutine with name VFIELD generates the values of the vector field V,
HOMOGENIZE determines the homogenized fitting $Y = \mathscr{H}_\epsilon(Z)$ of a curve representative Z.

We may achieve a good control over the evolution of the curves c^t when calling NEWCURVE and HOMOGENIZE equipped with suitable I/O routines independently from an interactive batch file. Instead of calculating Hausdorff distances of interpolating polygons of curve representatives it is convenient to use the approximate Hausdorff distance

$$\tilde{d}_H(Y, Z) = \max\left\{\max_{i=1}^{l}\min_{j=2}^{m} d(y_i, [z_{j-1}, z_j]),\right.$$
$$\left.\max_{j=1}^{m}\min_{i=2}^{l} d(z_j, [y_{i-1}, y_i])\right\} \tag{16}$$

between $Y = (y_1, \dots, y_l)$ and $Z = \{z_1, \dots, z_m\}$ where $d(a, [b, c])$ denotes the distance

$$\min_{0 \leqslant \xi \leqslant 1} \|a - (\xi b + (1-\xi)c)\|$$

of the point a from the line segment joining b and c.

The marked boxes in the flow-charts of Figs 1 and 2 show where parallel computation is suggested if parallel facilities are available.

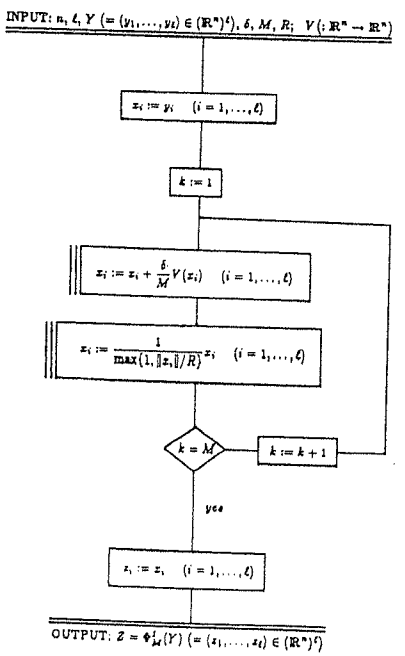

Fig. 1. Flow-chart of the program-segment NEWCURVE used in the procedure DDRP. (For notations see Section 2.)

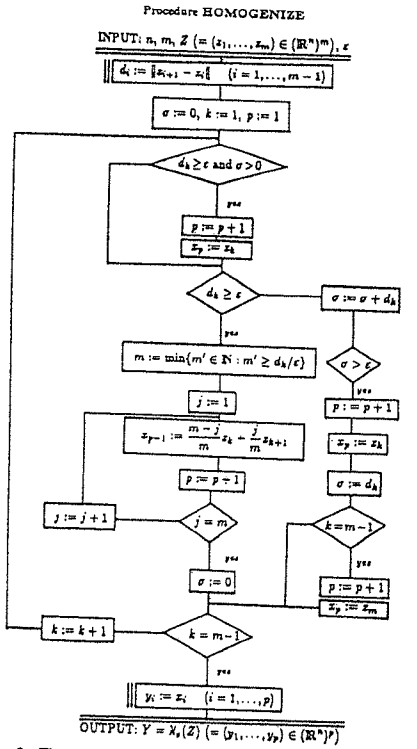

Fig. 2. Flow-chart of the program-segment HOMOGENIZE used in the procedure DDRP. (For notations see Section 2.)

3. FORTRAN REALIZATION

According to the previous section we divide the above described computational program into two subroutines, whose pertinent sections are the following:

NEWCURVE calculating the curve representative Z from Y in which a subroutine with name VFIELD generates the values of the vector field V, HOMOGENIZE determining the ϵ-homogenized fitting Y of a curve representative Z.

In sections of both subroutines the parameters are interpreted as follows:

N	dimension of the underlying space $\mathbb{R}^n$
T	corresponds to the time step δ
Y(I, J)	the Jth component of the Ith point of Y
Z(I, J)	the Jth component of the Ith point of Z
L	the number of points of both Y and Z in NEWCURVE
LX, LY	are the cardinalities of Y resp. Z in HOMOGENIZE
EPS, M, R	denote the controlling parameters ϵ, M, R.

Inputs of NEWCURVE are N, L, array Y (of type L × N), T, M, R.
Output of NEWCURVE is array Z (of type L × N).
Inputs of HOMOGENIZE are N, LZ, array Z (of type LZ × N), EPS.
Outputs of HOMOGENIZE are LY, array Y (of type LY × N).

For the sake of completeness we give also the FORTRAN code of the pertinent section of the subroutine HAUSD for determining the approximate Hausdorff distance $\bar{d}_H(Y, Z)$ of two curve representatives (consisting of LY, resp. LZ points in the Euclidean N-space).

Inputs of HAUSD are N, LY, LZ, array Y (of type LY × N), array Z (of type LZ × N).
Output of HAUSD is D.

```
      SUBROUTINE NEWCURVE(N,L,Y,Z,T,M,R)
      REAL Y(1000,30),Z(1000,30),X(30),V(30)
      ETA=T/M
      R2=R*R
          DO 7 I=1,L
          DO 1 J=1,N
    1     X(J)=Y(I,J)
          DO 5 K=1,M
          CALL VFIELD(N,X,V)
          DO 2 J=1,N
    2     X(J)=X(J)+ETA*V(J)
          D=0.
          DO 3 J=1,N
    3     D=D+X(J)*X(J)
          IF(D.GT.R2) THEN
          D=SQRT(D)
          DO 4 J=1,N
    4     X(J)=R*X(J)/D
          ENDIF
    5     CONTINUE
          DO 6 J=1,N
    6     Z(I,J)=X(J)
    7     CONTINUE
          RETURN
          END

      SUBROUTINE HOMOGENIZE(N,Y,LY,Z,LZ,EPS)
      REAL Y(1000,30),Y1(1000,30),Z(1000,30),D(1000)
      L=LZ-1
      DO 2 I=1,L
      S=0
      I1=I+1
      DO 1 J=1,N
    1 S=S+(Z(I1,J)-Z(I,J))**2
    2 D(I)=SQRT(S)
      DO 3 J=1,N
    3 Y1(1,J)=Z(1,J)
      LY1=1
      S=0.
      DO 8 I=1,L
      I1=I+1
      DD=D(I)
      IF((DD.GE.EPS).AND.(S.GT.0)) THEN
      LY1=LY1+1
      DO 4 J=1,N
    4 Y1(LY1,J)=Z(I,J)
      ENDIF
      IF(DD.GE.EPS) THEN
      M1=DD/EPS
      IF(DD.GT.M1*EPS) M1=M1+1
      DO 5 K=1,M1
      LY1=LY1+1
      DO 5 J=1,N
    5 Y1(LY1,J)=((M1-K)*Z(I,J)+K*Z(I1,J))/M1
      S=0.
      GOTO 8
      ENDIF
      S=S+DD
      IF(S.GT.EPS) THEN
      LY1=LY1+1
      DO 6 J=1,N
    6 Y1(LY1,J)=Z(I,J)
      S=DD
      ENDIF
      IF(I.EQ.L) THEN
      LY1=LY1+1
      DO 7 J=1,N
    7 Y1(LY1,J)=Z(I1,J)
      ENDIF
    8 CONTINUE
      DO 9 I=1,LY1
      DO 9 J=1,N
    9 Y(I,J)=Y1(I,J)
      LY=LY1
      RETURN
      END

      SUBROUTINE HAUSD(N,Y,LY,Z,LZ,D)
      REAL Y(1000,30),Z(1000,30)
      CALL HAUSD2(N,Y,LY,Z,LZ,D1)
      CALL HAUSD2(N,Z,LZ,Y,LY,D)
      IF(D1.GT.D) D=D1
      D=SQRT(D)
      RETURN
      END
```



```
      SUBROUTINE HAUSD2(N,G,LG,H,LH,D)
      REAL G(1000,30),H(1000,30)
      D=0.
      DO 4 J=1,LH
      DO 3 I=2,LG
      I1=I-1
      P=0.
      S1=0.
      DO 1 K=1,N
      P=P+(H(J,K)-G(I,K))*(G(I1,K)-G(I,K))
    1 S1=S1+(G(I,K)-G(I1,K))**2
      IF((P.LE.0.).OR.(S1.EQ.0.)) P=0.
      IF(S1.GT.0.) P=P/S1
      IF(P.GT.1.) P=1.
      S=0.
      DO 2 K=1,N
    2 S=S+(G(I,K)+P*(G(I1,K)-G(I,K))-H(J,K))**2
      IF((I.EQ.2).OR.(S.LT.D1)) D1=S
    3 IF(D1.LE.D) GOTO 4
      D=D1
    4 CONTINUE
      RETURN
      END
```

Program availability—The practical computational program as complete listings is available on request from the authors but will also be available on disk through circulation by the QCPE (Dömötör *et al.*, 1993a).

REFERENCES

Baker J. (1986) *J. Comput. Chem.* **7**, 355.
Baker J. (1987) *J. Comput. Chem.* **8**, 563.
Baker J. & Gill P. M. W. (1988) *J. Comput. Chem.* **9**, 465.
Bell S. & Crighton J. S. (1984) *J. Chem. Phys.* **80**, 2464.
Billing G. D. (1992) *Chem. Phys.* **161**, 245.
Chielke S. L. & Truhlar D. G. (1991) *J. Chem. Phys.* **95**, 5930.
Cummins P. L. & Gready J. E. (1989) *J. Comput. Chem.* **10**, 939.
Datta D. (1992) *J. Phys. Chem.* **96**, 2409.
Dömötör Gy., Stachó L. L. & Bán M. I. (1993a) QCPE program. Accepted.
Dömötör Gy., Stachó L. L. & Bán M. I. (1993b). In preparation.
Fukui K. (1970) *J. Phys. Chem.* **74**, 4161.
Fukui K., Kato S. & Fujimoto H. (1975) *J. Am. Chem. Soc.* **97**, 1.
Fukui K. (1981) *Acc. Chem. Res.* **14**, 363.
Gonzalez C. & Schlegel H. B. (1991) *J. Chem. Phys.* **95**, 5853.
Halgren T. A., Pepperberg I. M. & Lipscomb W. N. (1975) *J. Am. Chem. Soc.* **97**, 1248.
Halgren T. A. & Lipscomb W. N. (1977) *Chem. Phys. Lett.* **49**, 225.
Hirst D. M. (1985) *Potential Energy Surfaces: Molecular Structure and Reaction Dynamics.* Taylor & Francis, London.
Ischtwan J. & Collins M. A. (1991) *J. Chem. Phys.* **94**, 7084.
Ishida K., Morokuma K. & Komornicki A. (1977) *J. Chem. Phys.* **66**, 2153.
Jasien P. G. & Shepard R. (1988) *Int. J. Quant. Chem. Quant. Biol. Symp.* **22**, 183.
Jorgensen P. & Simons J. (1985) *Geometrical Derivatives of Energy Surfaces and Molecular Properties.* Reidel, Dordrecht.
Kliesch W., Schenk K., Heidrich D. & Dachsel H. (1988) *J. Comput. Chem.* **9**, 810.
Kraka E. & Dunning T. H. Jr (1990) In *Advances in Molecular Electronic Structure Theory*, Vol. 1, p. 129. JAI Press, Greenwich, Conn.
Lawley K. P. (1980) *Potential Energy Surfaces.* Wiley, New York.
McCullogh E. A. Jr & Silver D. M. (1975) *J. Chem. Phys.* **62**, 4050.
McIver J. W. & Komornicki A. (1974) *J. Am. Chem. Soc.* **96**, 5798.
Melissas V. S. & Truhlar D. G. (1992) *J. Chem. Phys.* **96**, 5758.
Mezey P. G., Peterson M. R. & Csizmadia I. G. (1977) *Can. J. Chem.* **55**, 2941.
Mezey P. G. (1987) *Potential Energy Hypersurfaces.* Elsevier, Amsterdam.
Mezey P. G. (1990) *J. Am. Chem. Soc.* **112**, 3791.
Muller K. & Brown L. D. (1979) *Theoret. Chim. Acta (Berl.)* **53**, 75.
Murrell J. N., Carter S., Farantos S. C., Huxley P. & Varandas A. J. C. (1984) *Molecular Potential Energy Functions.* Wiley, New York.
Natanson G. A., Garrett B. C., Truong T. N., Joseph T. & Truhlar D. G. (1991) *J. Chem. Phys.* **94**, 7875.
Peterson M. R. & Csizmadia I. G. (1977) *Prog. Theoret. Org. Chem.* **2**, 117.
Quapp W. & Heidrich D. (1984) *Theoret. Chim. Acta (Berl.)* **66**, 245.
Sana M., Reckinger G. & Leroy G. (1981) *Theoret. Chim. Acta (Berl.)* **58**, 145.
Scharfenberg P. (1980) *Theoret. Chim. Acta (Berl.)* **57**, 735.
Schlegel B. H. (1987) *Adv. Chem. Phys.* **67**, 249.
Silver D. M. (1966) *J. Chem. Phys.* **57**, 586.
Stachó L. L. & Bán M. I. (1992a) *Theoret. Chim. Acta (Berl.)* **83**, 433.
Stachó L. L. & Bán M. I. (1992b) *J. Math. Chem.* **11**, 405.
Stachó L. L. & Bán M. I. (1993) *Theoret. Chim. Acta (Berl.).* In press.
Tachibana A. & Fukui K. (1978) *Theoret. Chim. Acta (Berl.)* **49**, 321.
Tachibana A. & Fukui K. (1979) *Theoret. Chim. Acta (Berl.)* **51**, 189.
Truhlar D. G. (1981) *Potential Energy Surfaces and Dynamics Calculations.* Plenum, New York.
Xantheas S., Elbert S. T. & Ruedenberg K. (1991) *Theoret. Chim. Acta (Berl.)* **78**, 365.