Theoretical implications involved in the DDRP method *

Lászó L. Stachó

*Bolyai Institute for Mathematics, Attila József University, Aradi Vértanúk tere 1,
H-6720 Szeged, Hungary*

Miklós I. Bán

*Institute of Physical Chemistry, Attila József University, P.O. Box 105,
H-6701 Szeged, Hungary*

The aim of this paper is to prove that safe success in finding reaction paths (RPs) can only be expected from global path-determining methods. Some extensions of the mathematical arguments leading to the introduction of the DDRP (dynamically defined reaction path) method have been sketched. Four cases involving relaxation of analyticity, variability of the gradient field, minimum energy (reaction) paths (MEPs) and "golf pocket holes" on the potential energy surface (PES), and the rather strange consequences of the main theorem of the DDRP method giving a rigorous mathematical basis to chemical intuition in reaction kinetics have been discussed. The discussions show that the DDRP method – when changing the conditions and parameters – may, in essence, involve all other global methods. It has been shown that the DDRP method works in a stable way even for non-analytic though smooth energy functions; moreover, the gradient field can be replaced by other vector fields resulting in better convergence to the reaction path. As a by-product, the question of the existence of MEPs can safely be handled and golf pocket holes are constructed on the PES in order to prove that local methods have chance to search faithfully the RPs in complicated systems only if the energy function can be restored from its arbitrarily small pieces.

1. Introduction

Originally an *intrinsic reaction coordinate* (IRC) [1] was defined as a piecewise smooth curve on the potential energy surface (PES) joining two local minima whose tangent vector is always orthogonal (in mass-weighted coordinate system) to the equipotential contour lines of the PES. This means that, denoting the potential energy function by $U : \mathbb{R}^n \to \mathbb{R}$, a curve $c : [a, b] \to \mathbb{R}^n$ between two local minima a and b of U is an IRC if and only if

* This work was presented in parts at the 8th International Congress of Quantum Chemistry, Prague, Czech Republic, June 19–23, 1994; Addendum to the Book of Abstracts of the 8ICQC: P/I-129.

$$\frac{d}{ds}c(s) = \lambda(s)\nabla U(c(s)) \quad (s \in (a, b)) \tag{1}$$

for some scalar function $\lambda : (a, b) \to \mathbb{R}$. Notice that this concept is independent of parametrization and only depends on the level set structure of U. This latter observation has not yet been exploited on its merits.

In a series of papers [2–7] we have proposed a new curve variational method called the DDRP (dynamically defined reaction path) method which is numerically very stable in comparison with other global applications [8–13] of steepest descent path (SDP)-following methods. The theoretical basis of the DDRP method is the following mathematical fact (see Theorem 2.1 in ref. [2] and Theorem 2.2 in ref. [3]): If the potential function U is coercive and analytic with finitely many nondegenerate stationary points which are all of Siegel type then, for any piecewise analytic curve $c : [0, 1] \to \mathbb{R}^n$, the curves $[\exp(-t\nabla U)]c$ converge in Hausdorff distance to some IRC joining two stationary points. Moreover (see Proposition 3.10 in ref. [3]), there is a reparametrization $C^t : [0, 1] \to \mathbb{R}^n$ of the curves $[\exp(-t\nabla U)]c$ such that, for some finite sequence $0 = s_0 < s_1 < \ldots < s_N = 1$ of parameters, the points $C^t(s_i)$ converge to stationary points of U and all the derivatives of C^t converge locally uniformly to those of some IRC-piece in the intervals (s_i, s_{i+1}) as $t \to \infty$. It has already been illustrated by some practical examples [2,4] how this theorem can be used to achieve numerical procedures for calculating IRCs joining two given local minima of U even if the gradient field admits several bifurcations and the coercivity hypothesis fails. It can be conjectured that these methods work well even if the stationary points are degenerate or not of Siegel type although the convergence may be very slow in the former case.

In this paper we shall be concerned with some theoretical consequences of the mentioned theorem which give a rigorous mathematical background to some facts used intuitively in chemical practice and which may lead to several variants of the method.

2. Discussion

2.1. RELAXING ANALYTICITY

Formally the DDRP method can be applied to any potential function U with continuous gradient. Although the mathematically rather technical assumption of analyticity of U seems to be very natural even from the view point of a practicing chemist it may be important to know what happens if only a coercive potential function with continuous gradient is assumed and the requirement of analyticity is abandoned.

Let $U \in \mathcal{C}^1(\mathbb{R}^n)$ and suppose that $|U(x)|, \|\nabla U(x)\| \to \infty$ as $\|x\| \to \infty$ and the set $S := \{p : \nabla U(p) = 0\}$ of stationary points is finite. Consider any couple $p_1, p_2 \in S$

joined with a continuous curve $c : [0, 1] \rightarrow \mathbb{R}^n$ (i.e. $c(0) = p_1$, $c(1) = p_2$). Then the above conclusion of the theorem is still maintained in first order, i.e. the curves $[\exp(-t\nabla U)]c$ converge in Hausdorff distance to some IRC joining two stationary points. Now we sketch the proof of this fact.

Since the potential U is coercive there is some closed (and hence compact) region B containing the curve $C := c([0, 1])$ and the stationary points S such that

$$B^t := [\exp(-t\nabla U)]B \subset B \quad (t \geqslant 0) . \tag{2}$$

Then $B^{t_1} \subset B^{t_2}$ whenever $t_1 \geqslant t_2$. By the continuity of the exponential maps, each phase set B^t is compact. Hence the attractor $B^\infty := \bigcap_{t \geqslant 0} B^t$ is non-empty and

$$d_{\text{Hausdorff}}(B^t, B^\infty) \rightarrow 0 \quad (t \rightarrow \infty) ,$$

$$B^\infty = [\exp(-t\nabla U)]B^\infty \subset B \quad (-\infty < t < \infty) . \tag{3}$$

Example

We illustrate the above concepts by the model function $U_S := \Pi_{\epsilon,\eta=\pm 1}[(x - \epsilon)^2 + (y - \eta)^2] - (x - 1)^3 + (y - 1)^3$ used also in refs. [2] and [4]. In the present example we have chosen the square $\{(x, y) : |x|, |y| \leqslant 1.5\}$ for the starting region B^0. In figs. 1 and 2 one can see the phase regions B^t for $t = 0, 0.1, 0.3, 1, 2.5, 10, 30, 150$.

Let us consider any point $b \in B^\infty$ and write $b^t := [\exp(-t\nabla U)]b$ $(-\infty < t < \infty)$. Since $d/dt U(b^t) = \langle \nabla U(b^t), d/dt b^t \rangle = \langle \nabla U(b^t), -\nabla U(b^t) \rangle \leqslant 0$, the bounded functions $t \mapsto U(b^t)$ are decreasing on the whole real line. Thus any sequence $t_i \rightarrow \infty$
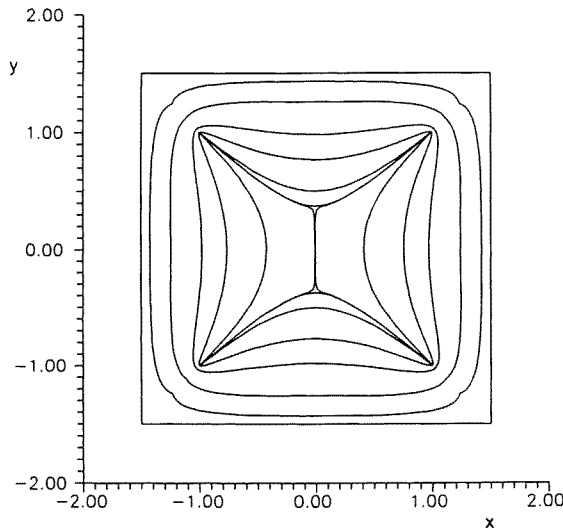

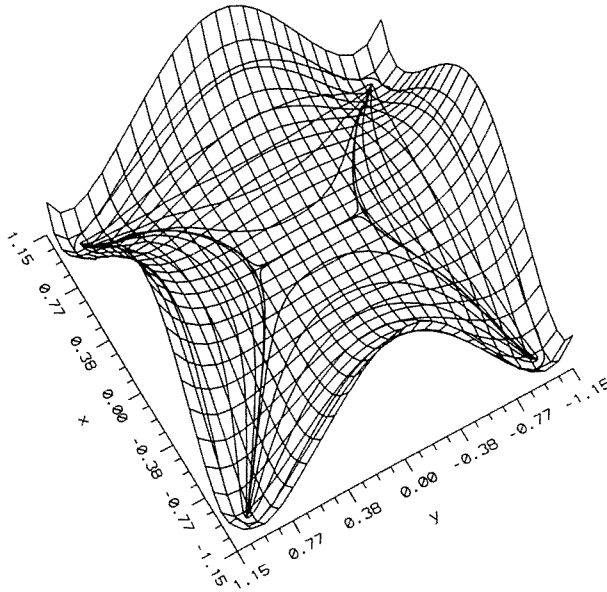Fig. 1. Phase regions B^t in the plane for the model function U_S, starting from B^0 (square).

Fig. 2. Phase regions of fig. 1 embedded in the PES of U_S.

(or $t_i \to -\infty$) admits a subsequence $(t_{i(k)})$ for which $(b^{t_{i(k)}})$ is convergent and $\|\nabla U(b^{t_{i(k)}})\| \to 0$. It follows that necessarily $\lim_{t \to \infty} b^t = b^\infty$ and $\lim_{t \to -\infty} b^t = b^{-\infty}$ for some stationary points t^∞, $t^{-\infty} \in S$ whenever $b \in B^\infty$. That is, B^∞ is the union of IRC pieces of U. By compactness arguments one can see that there are only finitely many of these IRC pieces, say $B^\infty = I_1, \ldots, I_N$. On the other hand, from the facts that for any $t \geqslant 0$ the curve $[\exp(-t\nabla U)]C$ joining p_1 with p_2 is contained in B^t and that $\lim_{t \to \infty} d_{\text{Hausdorff}}(B^t, B^\infty) = 0$, it follows that the limit set

$$I^* := \{p : \exists(t_i, s_i)_{i=1,2,\ldots} \ t_i \to \infty, 0 \leqslant s_i \leqslant 1, \lim_i [\exp(-t\nabla U)]c(s_i) = p\} \qquad (4)$$

is compact and tangent to the gradient field of U, i.e.

$$[\exp(-t\nabla U)]I^* \subset I^* \quad (-\infty < t < \infty). \qquad (5)$$

Furthermore, since each curve $[\exp(-t\nabla U)]C$ $(t \geqslant 0)$ is connected, the set I^* is also connected. Hence

$$I^* = I_{n_1} \cup \ldots \cup I_{n_l},$$

$$d_{\text{Hausdorff}}([\exp(-t\nabla U)]c, I^*) \to 0 \quad (t \to \infty) \qquad (6)$$

for some consecutive subfamily $I_{n_1}, \ldots, I_{n_l}$ of IRC pieces of U joining p_1 and p_2, thus completing the proof.

2.2. NEW VARIANTS OBTAINED BY TRANSFORMING THE GRADIENT FIELD

Notice that the virtual phases

$$\{[\exp(-t\nabla U)]p : t \geqslant 0\} \quad (p \in \mathbb{R}^n \text{ fixed}) \tag{7}$$

are invariant if we replace ∇U with any continuous vector field of the type

$$V(p) := \lambda(p) \cdot \nabla U(p) \quad \lambda(p) > 0 \text{ if } p \notin S, \quad V(S) = 0. \tag{8}$$

We have already used this idea in a special case [7,14] where we assumed the replacement of the potential function U with a function of the form $g(U)$ ($g : \mathbb{R} \to \mathbb{R}$) strictly increasing) may accelerate path-following. Knowing that the DDRP method is effective also in the case of a $\mathcal{C}^1$-smooth potential U, one can show, by a compactness argument, the following more general fact.

Let again $U \in \mathcal{C}^1(\mathbb{R}^n)$ and $|U(x)|$, $\|\nabla U(x)\| \to \infty$ as $\|x\| \to \infty$ with finite $S := \{p : \nabla U(p) = 0\}$. If the function $\lambda \in C(\mathbb{R}^n)$ is such that $\lambda(p) > 0$ for $p \in \mathbb{R}^n$ and $c : [0, 1] \to \mathbb{R}^n$ is a continuous curve joining the stationary points $p_1, p_2 \in S$ then the curves

$$c^t := [\exp(-\lambda\nabla U)]c \quad (t \geqslant 0) \tag{9}$$

converge uniformly to some reaction path (RP) of U joining p_1 with p_2.

Example

The function $U_M := xy^2 - x^3/3 + (x^2 + y^2)^2/4$ has a stationary point of order 2 (i.e. a so-called monkey saddle) at the origin. It also has three minima at the points $(0, 1)$, $(-1/2, \pm\sqrt{3}/2)$. The system of RPs of U consists obviously of the three straight line segments joining the origin with the minima (see figs. 3 and 4). Since the function U is very flat around the origin (due to the stationarity of 2nd order) the convergence of any RP-following method based upon the direct calculation of $-\nabla U$ becomes extremely slow. However, if we replace the vector field $-\nabla U$ with the vector field $\mathcal{V} := -\|\nabla U\|^{-1/2}\nabla U$ then the searching will be performed approximately with a speed experienced in neighbourhoods of saddle points of 1st order. In figs. 5–8, starting from the triangle C^0 with vertices at the minima, we compare a conventional application of the DDRP method (i.e. calculating the curves $C^t := [\exp(-t\nabla U)]C^0$) with the application using the field $\mathcal{V}$ instead of $-\nabla U$ (i.e. calculating the curves $K^t := [\exp(-t\mathcal{V})]C^0$). In figs. 5 and 6 the curves C^t and in figs. 7 and 8 the curves K^t are shown for $t = 0, 15, 30, 60, 170, 350$. It is worth noticing that the components of $-\nabla U$ may vanish in the computer representation in some small neighbourhood of the origin while $\mathcal{V}$ is not numerically zero at most points of the same neighbourhood. (Naturally, fine numerical methods are necessary to obtain exact values for the components of $\mathcal{V}$ if ∇U vanishes numerically.)
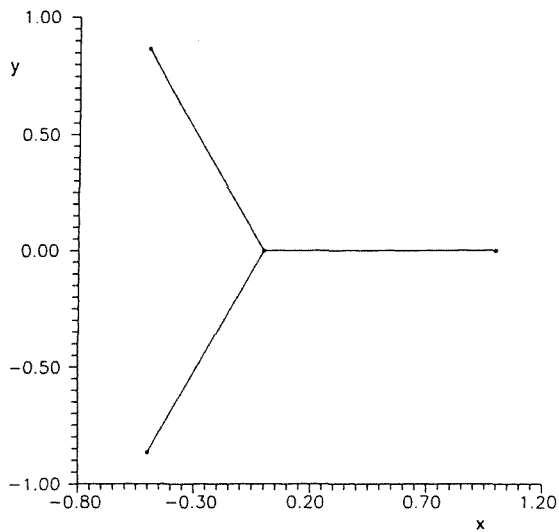
Fig. 3. Complete IRC system of the monkey saddle function U_M in the plane.
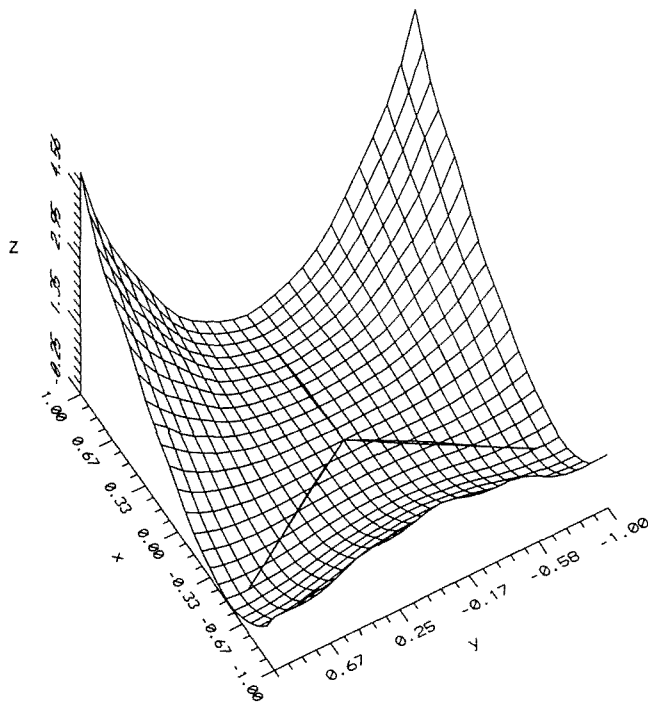

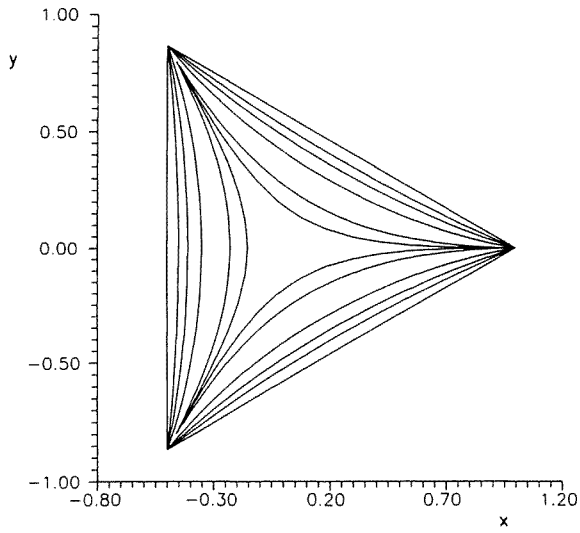Fig. 4. IRCs of fig. 3 embedded in the PES of U_M.

Fig. 5. Phase curves C^t in the plane, calculated by the conventional DDRP method, up to the virtual time $t = 350$.
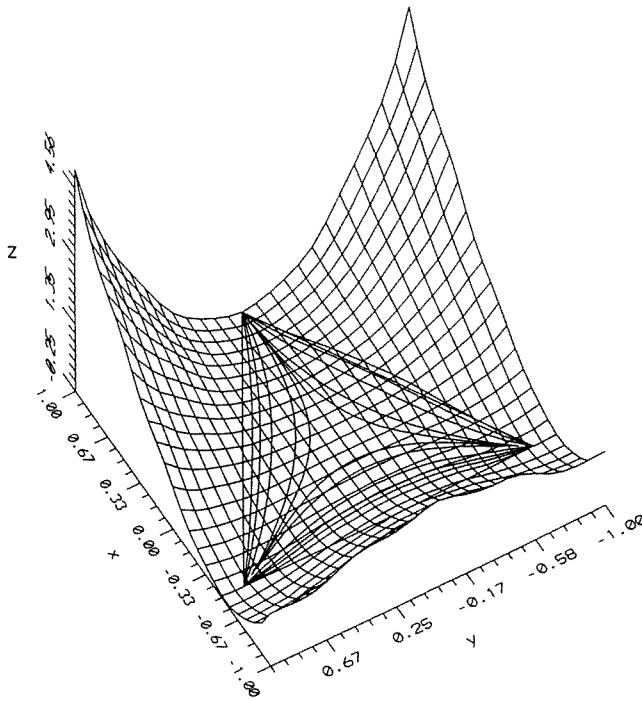

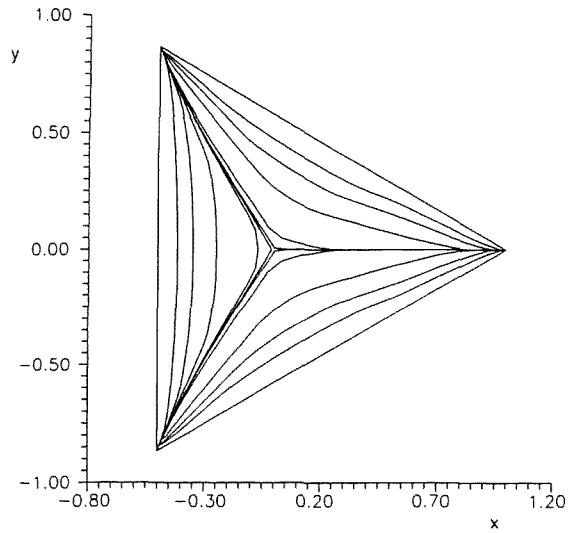Fig. 6. Phase curves C^t of fig. 5 embedded in the PES of U_M.

Fig. 7. Phase curves K^t in the plane, calculated by a square-root modified DDRP method, up to the same virtual time $t = 350$ as in fig. 5.
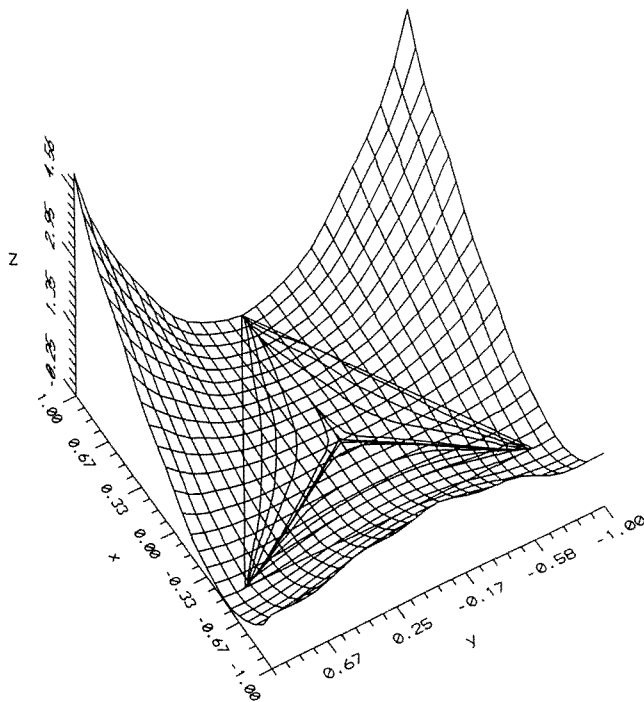

Fig. 8. Phase curves K^t of fig. 7 embedded in the PES of U_M.

Remark

It seems all known global curve variational methods have the following common feature. They start from some initial curve C^0 in the configuration space (where the energy function U is defined). The only information they use to transform C^0 to some RP is involved in the gradient field of U, therefore, the obtained approximation curves

$$A^\tau \quad (\tau \to \infty, A^0 = C^0) \tag{10}$$

should always be the subsets of the 2-dimensional (2D) surface

$$S := \bigcup_{t \geqslant 0} [\exp(-t\nabla U)]C^0$$

$$\{c(s,t) : a \leqslant s \leqslant b, t \geqslant 0\} \tag{11}$$

if C is parametrized as $C = \{c(s,0) : a \leqslant s \leqslant b\}$ and $c(s,t) := [\exp(-t\nabla U)]c(s,0)$. To our argumentations in section 2.1, there exists a finite set $I_1, \ldots, I_N$ of SDPs or steepest ascent paths (SAPs) of U such that

$$d_{\text{Hausdorff}}\left(\{c(t,s) : a \leqslant s \leqslant b, t \geqslant \theta\}, I_1 \cup \ldots \cup I_N\right) \to 0 \quad (\theta \to \infty). \tag{12}$$

Let us now define

$$\alpha(\tau) := \sup\{\theta : A^\tau \subset c([a,b] \times [\theta, \infty))\} \tag{13}$$

for every index τ. It follows that

$$d_{\text{Hausdorff}}(A^\tau, I_1 \cup \ldots \cup I_N) \to 0 \quad \text{if and only if} \quad \alpha(\tau) \to \infty \quad \text{for } \tau \to \infty. \tag{14}$$

Given a starting curve C^0, it is a very important task to find methods whose lower parameter measuring function $\alpha(\theta)$ tends to ∞ as rapidly as possible. However, numerical experiences establish that methods with too ambitious accelerations jeopardize the convergence and numerical stability seriously. E.g. Liotard's chain method [8] seems to try defining the curves $A^\tau := \{c(s, \tau) : a \leqslant s \leqslant b, \tau \geqslant 0\}$ (so the same curves as the simple variant of the DDRP method) by trying to determine points of $A^{\tau+\delta}$ in orthogonal directions in S to the tangent of A^τ. If the parametrization $c(s, t)$ happens to be very smooth with small curvatures, even relatively large values of δ may lead to satisfactory new approximations. However, this is not the case if, e.g., the system $I_1, \ldots, I_N$ bifurcates and then there is not known any good estimate by the aid of which one can keep the step parameters of the chain method within numerically reliable limits without requiring superfluously large calculations.

The fact that the methods previously considered explore some *only 2D* surface (the surface S in the above subsection) independently of the dimension of the configuration space is extremely important. This implies namely that the numerical

costs of DDRP methods do not increase quickly with the dimension of the configuration space. Even for the simple DDRP method we can give the following rough comparison with local (sequential) RP-following methods. If each of the curves $C^t := \{c(s, t) : a \leqslant s \leqslant b\}$ has a length at most λ and it can be reliably represented by a sequence of discrete points lying in a distance $\approx \epsilon$ consecutively then the surface S is explored by $\approx \lambda/\epsilon$ subcurves each of which is generated in an analogous manner to sequential methods. Therefore, even in worst cases, the DDRP method finds the RP system $I_1 \cup \ldots \cup I_N$ at most $\approx \lambda/\epsilon$ times longer time in a sequential computer than a local method passes through $I_1 \cup \ldots I_N$ in its optimum functioning. *Thus if $a \approx \lambda/\epsilon$-fold parallellizability is available, the DDRP method requires the same computer time as a local method does.* Finally we remark that it seems to be reasonable to conjecture that the value of λ/ϵ increases in the order of the square root of the dimension of the configuration space.

2.3. ON MINIMUM ENERGY (REACTION) PATHS

For a long time studies focused on cases where the PES had the structure of a canyon. In such investigations the potential energy function U was assumed to be differential-topologically equivalent to $x \mapsto \|x - e\|^2 \|x + e\|^2$ where $e \in \mathbb{R}^n$ is any unit vector. The behaviour of this typical model function suggested several alternative RP definitions which coincide with the IRC in the simplest cases. The concept of minimum energy (reaction) path (MEP) is particularly worth mentioning. By definition MEP is a piecewise smooth curve joining two local minima on the PES such that the potential energy is locally minimal at any point of the curve along the hypersurface lying orthogonally to the tangent vector of the curve and passing through the point in question. It is easy to see that MEPs are becoming automatically IRCs if the typical sufficient condition

$$\langle H(c(s))v, v \rangle \geqslant 0 \quad (s \in (\alpha, \beta), v \perp \frac{d}{ds} c(s)) \tag{15}$$

is realized for $c : [\alpha, \beta] \to \mathbb{R}^n$. Here H denotes the Hessian (second derivative matrix) of U and $v \in \mathbb{R}^n$.

It would be interesting theoretically to find reasonable conditions for the potential function admitting MEPs between any couple of local minima. Nevertheless, a 2-variable pure mathematical counterexample contradicting immediately this ambition can very easily be constructed. Namely, a function $U : \mathbb{R}^2 \to \mathbb{R}$ whose graph is the surface of a golf area with a pocket hole on the top of a hill would be adequate for this purpose. Indeed, any curve leading from the bottom of this pocket hole to another hole anywhere on the PES passes through the convex surface piece around the top, thus, in this case, the sectional second derivative $\langle H(c(s))v, v \rangle < 0$ in each direction v. As far as we know there are no chemical examples for the higher dimensional analogy of the golf pocket hole. However, the conformal changes of the catechol molecule discussed by Mezey [15] give a 2D section of the potential

function which is differential-topologically equivalent to the function with an H-shaped family of SDPs between stationary points discussed in our previous papers [2,4]. Here we have three consecutive saddle points (SPs) along the RP joining two diametrically opposit minima, and this is the only way of connecting them by a simple curve consisting of a family of SDPs. This RP is no MEP because the starting direction of the second segment (leading from one of the outer SPs to the SP at the origin) is parallel to the eigenvector with negative eigenvalue of the Hessian. Notice that (in this case) a conventional method [16–18] following stepwise a SAP starting from a local minimum would find first a SP (as classically expected), but then the SDP from this SP would lead to a non-opposit minimum. Such searchings may be carried out also by MEP-methods as those of Schlegel [19]. The right solution for continuing the first SAP is to issue again a SAP from this first SP in the negative eigendirection of the Hessian (toward zero). Now we can see that this can only be successful if we use a searching for SA directions. No MEP-methods can perform this task since the first segment of this RP-piece is no more a MEP-piece (now). Having reached the SP at the origin one can continue in a similar fashion except that now by searching for SDPs instead of SAPs. The numerically most stable non curve-variational methods suggest to locate first a SP with a unique negative eigenvalue of the Hessian and then issuing SDPs to both negative eigendirections. This idea goes back to that of Murrell and Laidler [20] who proved that the energy function U assumes its maximal value along a path with minimal energy increment in a SP between two minima, with a unique negative Hessian eigenvalue. Nevertheless, their remarkable argument does not contain any information concerning the existence of RP of minimal energy increment. The following mathematical example suggests that some kind of coercivity of U is necessary for the existence of such RPs. Let $U : \mathbb{R}^3 \to \mathbb{R}$ be the function

$$U := U_0 \cdot P_1 \cdot P_2, \quad \text{where}$$

$$U(x, y_1, y_2) := \exp((x^2 - y_1^2 - y_2^2)/4),$$

$$P_k(x) := 1 - \exp(-1000 \cdot \|x - e^{(k)}\|), \quad e^{(k)} := (0, 0, (-1)^k) \quad (k = 1, 2).$$

$$(16)$$

Thus the graph of the function U will be approximately that of U_0 equipped with two golf pocket holes at the points $e^{(1)}$, $e^{(2)}$. The straight line segment $\{(1-t)e^{(1)} + te^{(2)} : 0 \leqslant t \leqslant 1\}$ is the only RP joining the minima $e^{(1)}$, $e^{(2)}$ with a curve passing through the origin where $U(0) \approx 1 (> 0.99)$. On the other hand, the function U assumes values strictly $< U(0)$ along any curve c between $e^{(1)}$ and $e^{(2)}$ and passing completely in the plane $Y := \{(0, y_1, y_2): y_1, y_2 \in \mathbb{R}\}$. Observe that the projection $\Pi : (x, y_1, y_2) \mapsto (0, y_1, y_2)$ decreases the values of U (i.e. $U(\Pi x) \leqslant U(x)$). Therefore, when looking for curves with a minimal energy increment between $e^{(1)}$ and $e^{(2)}$, we may restrict our attention to the plane Y. Nevertheless, the function

$(y_1, y_2) \mapsto \exp(-(y_1^2 + y_2^2)/4)$ has a negative definite Hessian for $y_1^2 + y_2^2 < 2$ and hence U on Y behaves around the golf pocket bottoms $e^{(1)}$, $e^{(2)}$ similarly as in the hill-top case. Therefore minimum energy increment paths joining $e^{(1)}$ with $e^{(2)}$ cannot be MEPs. However, if we allow ideal points in infinity (in projective sense) we may really find RPs in Y with a minimal energy increment between $e^{(1)}$ and $e^{(2)}$. These are RPs passing very closely to the straight lines $x = 0$, $y_2 = \pm 1$ through the lowest points of the brinks of the two golf pocket holes (which are really SPs with one negative and two positive Hessian eigenvalues). A practical way of avoiding infinite regions in calculations is to make the function U coercive by a perturbation outside some relevant domain. We may add, e.g., the function $\phi(x) := (\|x\|/10)^{10}$ to U by which numerically almost no influence has been made on U in case of $\|x\| \leqslant 6$ apart from making $U + \phi$, along with its gradient, to tend rapidly to ∞ if $\|x\| \to \infty$. In section 2.1 we have established the global existence of RPs between any couple of local minima for coercive analytic energy functions U with finitely many stationary points each of which being of first order. The proof ensures that the phase flow $[\exp(-t\nabla U)]x$ of the negative gradient of the energy function U guides any polygon joining two given stationary points, which intersects finitely many times the boundaries of the catchment regions [21] of the stationary points of U, to some RP [1]. By the aid of our mathematical theorem we can immediately show the existence of curves with minimal energy increment which are RPs between any pair of stationary points of an energy function U satisfying (the not too restrictive) hypothesis of the theorem.

Indeed, let a and b be two stationary points of U and set $u_0 := \inf_c \max U(c)$ for all curves c joining a with b. Given any $\epsilon > 0$, we can find a polygon Z_ϵ between a and b such that $\max U(Z_\epsilon) < u_0 + \epsilon$. By the theorem the curves $[\exp(-t\nabla U)](Z_\epsilon)$ converge uniformly to some RP I_ϵ when parametrized arclength proportionally on $[0, 1]$. The energy values decrease along the phase flow of its negative gradient, therefore $\max U(I_\epsilon) \leqslant \max U(Z_\epsilon)$ for every $\epsilon > 0$. Since an RP is the union of a finite family of SDPs/SAPs, the energy function U takes its maximum along I_ϵ at some stationary point p_ϵ^*. By assumption we have only finitely many stationary points. Hence some sequences $p_{\epsilon_n}^*$ ($n = 1, 2, \ldots$) with $\epsilon_n \to 0$ consist of all the same point say p^*. It follows that

$$\max U(I_{\epsilon_1}) = U(p^*) = \max U(I_{\epsilon_n}) \leqslant u_0, \tag{17}$$

proving that the RP I_{ϵ_1} is a curve with minimal energy increment between a and b. As a corollary from Murrell–Laidler's argument [20] one can obtain that the stationary point p^* with maximal energy value on I_{ϵ_1} should have a unique negative Hessian eigenvalue. We emphasize the assumptions concerning the stationary points, i.e. in particular that they should be all of first order. Some authors claim that this latter condition is satisfied automatically in chemical situations because paths with minimal energy increment should always exist. This statement seems to

---

[1] Given any curve we can find such polygons arbitrarily close to it with the same endpoints.

be unproved. Note that every energy function with finitely many stationary points can be perturbed analytically in such a way that it can be kept arbitrarily close to the original energy function outside a given neighbourhood of its stationary points and although the stationary points of the perturbed function are the same they will be of higher order. Indeed, if $p_1, \ldots, p_m$ are the stationary points of U then the composite function $U \circ \Psi_{1,\epsilon} \circ \cdots \circ \Psi_{m,\epsilon}$ (where $\Psi_{i,\epsilon} : x \mapsto p_i + (1 - \exp(\|x - p_i\|/\epsilon) \cdot (x - p_i))$ has the required properties for sufficiently small $\epsilon > 0$. Unfortunately an inverse construction cannot be carried out in general without changing the number of stationary points. However, if we are interested only in a uniform approximation then some polynomial function with finitely many stationary points and sufficiently close to a given compact region is always available as a consequence of the classical Stone–Weierstrass's approximation theorems [22].

2.4. GOLF POCKET HOLES ON THE PES

We can use the golf pocket hole construction of the above subsection to transform a given potential function U into some $\tilde{U}$ that differs from U only on given small open sets in which it has new stationary points. If $U > 0$ (this may be assumed without loss of generality in our context) then

$$\tilde{U}(x) := U(x) \prod_{i=1}^{N} \Phi\left(\frac{1}{\epsilon}(x - p_i)\right) \quad (x \in \mathbb{R}^n) \tag{18}$$

differs from U in the neighbourhoods of radius $\epsilon > 0$ of the points $p_1, \ldots, p_N$ which are its new stationary (actually minimum) points where $\Phi \in \mathbb{C}^\infty(\mathbb{R}^n)$ is the (non-analytic) golf pocket hole function

$$\Phi(x) := \exp(-(1 - \|x\|)^{-2}) \quad \text{for } \|x\| < 1, \quad \Phi(x) := 1 \quad \text{for } \|x\| \geqslant 1. \tag{19}$$

We can use this remark to show that the DDRP method is suitable in a globally very accurate and steady gradient following. Given an arbitrary point a we can generate a meta-IRC issued from a as follows. Let p be the stationary point whose catchment region contains the point a, and let $c : [0, 1] \to \mathbb{R}^n$ be any polygon joining a with p. Then making a sufficiently small golf pocket hole around a on the energy surface, one piece of the resulting IRC of the (by the golf pocket hole) perturbed energy function obtained by the DDRP method is a small perturbation of the meta-IRC we are looking for. In practice we need not perturb the energy function but we have to add the point a to the resulting curve after each homogenization as the starting point of the new phase curve. Moreover, if we are satisfied with lower stability then it is not necessary to know even the attracting stationary point p. We may then start the algorithm with the modified homogenization from the trivial curve $[0, 1] \to \{a\}$. In this case the algorithm works as a corrector method. From the previous considerations we can also see that by adding a given couple of points to the homogenized phase curves as new starting and end points, respectively, we

achieve in limit an IRC with two meta-IRC pieces connecting the two given points. Even though the energy surface does not satisfy our rigorous hypothesis, the above gradient following can be applied to find such valley bottom curves as the sinuous curve in Schlegel's surface [23].

Example

It is well-known that for a generic positive definite quadratic energy function of two variables the axis of its eigenvectors corresponding to the higher eigenvalue is a meta-IRC while all other meta-IRCs start from the origin in the direction of the axis formed by the eigenvectors corresponding to the lower eigenvalue. Since through every point, except the origin, a unique meta-IRC passes, a local method cannot decide which of them should be followed if we only know that the IRC starts in the direction of some eigenvector corresponding to the lower eigenvalue. Indeed, we may multiply U by, e.g., a golf pocket hole function

$$G_{(a,b)} := \phi((x-a)^2 + (y-b)^2) \quad (x, y \in \mathbb{R}), \tag{20}$$

where $\phi : \mathbb{R} \rightarrow [0, 1]$ is an increasing C^∞-smooth function such that

$$\phi(r) = 0 \quad \text{for } r \leqslant 0, \quad \phi(r) = 1 \quad \text{for } r \geqslant 0.01. \tag{21}$$

Then the modified function $G_{(a,b)}$ has a golf pocket hole of radius 0.1 around the point (a, b) and it coincides with U outside the golf pocket hole. In figs. 9 and 10 we have considered the function $(x^2 + 4y^2)G_{(1,0)}G_{(1/2,\sqrt{3}/4)}G_{(\sqrt{3}/2,1/4)}G_{0,1/2}$. Here each golf pocket hole bottom is joined to the origin by an IRC piece which practically does not differ from the meta-IRCs between these points and the origin.

3. Conclusions

In section 2.1 we have shown that the DDRP method possesses all the properties and abilities of other gradient following methods, even without complying with
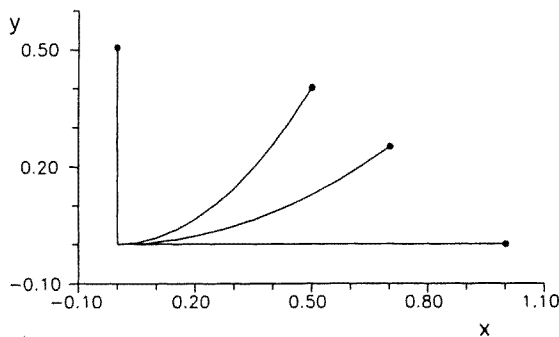

Fig. 9. IRCs joining the golf pocket holes on the PES $z = x^2 + 4y^2$ projected onto the plane.
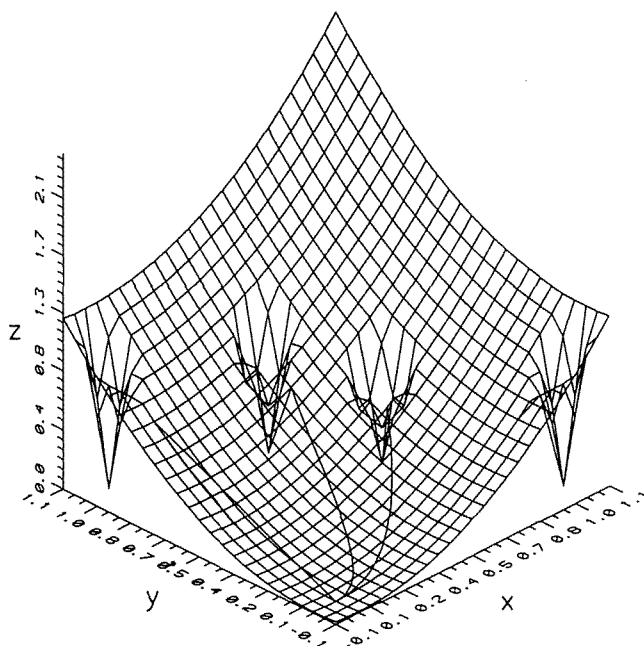
Fig. 10. PES $z = x^2 + 4y^2$ modified by the golf pocket holes showing the embedded IRCs.

the requirement of analyticity of the energy function. However, although the DDRP method can still be used if analyticity is not fulfilled, the RP cannot generally be approximated with the same theoretical accuracy in higher order as in the analytical case and we may count with some even stranger consequences if the energy function is not analytic. In section 2.2 the general theoretical conditions of convergence acceleration have been discussed. It seems that all known global methods different from ours try to approximate phase curves in ways which could be obtained by some version of the modified DDRP algorithm described in ref. [4]. Section 2.3 presents a mathematically complete formulation of the Murrell–Laidler argument [20]. Although their proof can be considered almost perfect, it did not deal with existence problems. The DDRP method supplies the missing links. The study of golf pocket holes in section 2.4 serves to point out essential disadvantages of local methods with respect to global ones. Namely, if the energy function cannot be reconstructed from its arbitrarily small piece as is the case with, e.g., Taylor series of analytic functions (this may happen even to functions differentiable infinitely many times), then local RP-following [19,23] can be misleading. Unfortunately, such circumstances have usually not been treated by rigorous theoretical thoroughness, and to our knowledge, the possibility that the occurrences of $\mathcal{C}^\infty$-smooth perturbations may reorganize the structure of reaction surface has only been investigated in this present article. Note that the golf-hole construction is perhaps only the simplest way to produce counterexamples, and more sophisticated

non-analytic constructions can be given whose chemical relevance is not excluded. Numerical calculations with piecewise different analytic formulas (typical iterations using different lengths) may produce an effect similar to non-analytic perturbations. It seems some *pars pro toto* principle is required in order that the PES could be searched reliably by local methods. The early successes of local methods may indicate that for real chemical energy functions the *pars pro toto* principle actually prevails. However, this fact, to our knowledge, has not been proved although it would be very important to have it done. Owing to the 2D nature of the test functions investigated so far, it can also be conceived that the golf-hole paradox discussed above cannot occur at all if starting from a SP. This is, nevertheless, strictly the peculiarity of 2D cases. If a golf hole exists, e.g., at some place of the PES, from that point a new RP could always be issued toward a SP and there is no reason why this new RP could be followed by any conventional local method, while we can prescribe for most global methods that a RP, joining the bottom of the golf hole with the SP, should be obtained. After all, we can state that our global DDRP method is superior even to other global gradient following methods, at least in its rigorous mathematical foundation, its high stability, versatility and easy parallelizability.

It must be emphasized that the above conclusions have been drawn on the basis of mere mathematical considerations. Nevertheless, it is fairly obvious to assume that the chemical PESs also comply with appropriate theoretical treatments based on substantiated mathematical theorems, therefore, the "implications involved in the DDRP method" are hopefully not the results of abstract reasoning solely.

Acknowledgement

References

[1] K. Fukui, J. Phys. Chem. 74 (1970) 4161.
[2] L.L. Stachó and M.I. Bán, Theor. Chim. Acta 83 (1992) 433.
[3] L.L. Stachó and M.I. Bán, J. Math. Chem. 11 (1992) 405.
[4] L.L. Stachó and M.I. Bán, Theor. Chim. Acta 84 (1993) 535.
[5] L.L. Stachó and M.I. Bán, Comp. Chem. 17 (1993) 21.
[6] Gy. Dömötör, M.I. Bán and L.L. Stachó, J. Comput. Chem. 14 (1993) 1491.
[7] M.I. Bán, Gy. Dömötör and L.L. Stachó, J. Mol. Struct. (Theochem) 311 (1994) 29.
[8] D.A. Liotard and J.-P. Penot, in: *Numerical Methods in the Study of Crytical Phenomena*, eds. J. Della Dora, J. Demongeot and B. Lacolle (Springer, Berlin, 1981) p. 213.
[9] L.R. Pratt, J. Chem. Phys. 85 (1986) 5045.
[10] R. Elber and M. Karplus, Chem. Phys. Lett. 139 (1987) 375.

[11] A. Ulitsky and R. Elber, J. Chem. Phys. 92 (1990) 1510.

[12] S. Fischer and M. Karplus, Chem. Phys. Lett. 194 (1992) 252.

[13] D.A. Liotard, Int. J. Quant. Chem. 43 (1992) 723.

[14] L.L. Stachó, Gy. Dömötör and M.I. Bán, presented at The Royal Society of Chemistry Faraday Division, Symp. No. 29, Oxford, UK (1993); J. Chem. Soc. Faraday Trans. 90 (1994) i, to be published in J. Mol. Struct. (Theochem).

[15] P.G. Mezey, J. Am. Chem. Soc. 112 (1990) 3791.

[16] e.g. J.W. McIver Jr. and A. Komornicki, J. Am. Chem. Soc. 94 (1972) 2625; ibid 96 (1974) 5798.

[17] I. Bálint and M.I. Bán, Theor. Chim. Acta 63 (1983) 255;
I. Bálint and M.I. Bán, Int. J. Quant. Chem. 24 (1983) 161 and ibid 25 (1984) 667; and references therein.

[18] K. Ruedenberg and J.Q. Sun, J. Chem. Phys. 100 (1994) 5836.

[19] C. Gonzales and H.B. Schlegel, J. Phys. Chem. 94 (1990) 5523; J. Chem. Phys. 90 (1989) 2154 and ibid 95 (1991) 5853.

[20] J.N. Murrell and K.J. Laidler, Trans. Faraday Soc. 64 (1968) 371.

[21] e.g. P.G. Mezey, Theor. Chim. Acta 58 (1981) 309.

[22] e.g. L.H. Loomis, *An Introduction to Abstract Harmonic Analysis* (Van Nostrand, Toronto, 1953).

[23] H.B. Schlegel, J. Chem. Soc. Faraday Trans. 90 (1994) 1569.