

Reaction pathways and convexity of the potential energy surface: application of Newton trajectories

Michael Hirsch and Wolfgang Quapp*

Mathematical Institute, University of Leipzig, Augustus-Platz, D-04109 Leipzig, Germany
E-mail: quapp@rz.uni-leipzig.de

Received 25 June 2004; revised 5 July 2004

The reaction path is an important concept of theoretical chemistry. We employ the definitions of the intrinsic reaction coordinate (IRC), the gradient extremal (GE), and the Newton trajectory (NT). The usual imagination in chemistry is that a minimum energy path is in a convex region of the potential energy surface. We describe different schemes of convexity to handle the situation. It comes out that NTs are the best ansatz for the problem: NTs, which monotonically increase (or monotonically decrease), are automatically strictly pseudo-convex throughout, and they go throughout along a valley between minimum and saddle point.

KEY WORDS: potential energy surface, reaction path, projected gradient, IRC, Newton trajectory, pseudo-convexity

AMS subject classification: Primary: 58C15, 26B25
Secondary: 53A04, 58K05, 52A20, 53Z05

1. Introduction

The concept of the minimum energy path (MEP) or reaction path (RP) of an adiabatic potential energy surface (PES) is the usual approach to the theoretical kinetics of larger chemical systems [1,2]. It is roughly defined as a line in coordinate space, which connects two minima by passing the saddle point (SP), the transition structure of a PES. The energy of the SP is assumed to be the highest value tracing along the RP. It is the minimal energy a reaction needs to take place.

Reaction theories are based either implicitly (transition state theory), or explicitly (variational transition state theory) on the knowledge of the RP [2]. These theories require local information about the PES along the RP only. They circumvent the dimension problem for medium-sized or large molecules: it is impossible to fully calculate their PES.

* Corresponding author.

The starting point is a geometrically defined pathway which may serve as an RP. Geometrically defined means that only properties of the PES are taken into account, and that no dynamic behavior of the molecule is taken into consideration. Any parameterization s of the RP $\mathbf{x}(s) = (x^1(s), \dots, x^n(s))^T$ is called reaction coordinate. How an RP ascends to the SP is an uncertainty of the RP definition. We use here the IRC [3] and the distinguished or driven coordinate method [4,5] in the modern form of RGF [6,7], also called Newton trajectory (NT). We insist that the search for an appropriate MEP is not necessarily equivalent to the finding of the steepest descent (SD) pathway from the SP. It is not obvious that the SD is the best choice to describe kinetics – in general.

Usually, in one's imagination the MEP is situated in a valley of the PES. The different forms of an RP should connect minimum and SP of index one going through a valley. However, it contradicts the examples where the IRC does not fulfill the property: it is known that the IRC can go over a ridge of the PES, cf.[8]. NTs can also go over a ridge of the PES, then they have a turning point (TP) [4,9,10]. Consequently, the classification of IRC and NTs belonging to MEPs or not, is of interest.

In this paper we show that the IRC is an MEP if it does not cross the pseudo-convexity boundary of the PES, which is defined by a simple formula [11]. For NTs, the TP-case divides them into those which can serve as loose RPs, and others: if the NT does not contain a TP at the pathway from minimum up to the SP of index one, it can be used as an RP model [12]. Additionally, we show that NTs have a nice property: if they monotonically increase from minimum to SP then they automatically take a course throughout in a valley [11]. To our knowledge, NTs are the sole curves with this property.

The paper is organized as follows. Sections 2–4 repeat known properties of PESs and their inflection points, of steepest descent on the PES, and of projections. With Section 5 the main part begins giving some properties of Newton trajectories known up to date. Section 6 starts with the definition of the pseudo-convexity of a PES and the relation of NTs to that important property. It finishes with three theorems, being the main content of this paper, to NTs and valley structures of the PES. In Section 7 we discuss the results and give some conclusions. An appendix contains the used formulas of the model PESs.

2. Potential energy surface

The adiabatic PES of the molecular system of observation is the basis of our treatment. Using the Born–Oppenheimer approximation, we assume that the movement of the electrons and of the atom kernels are decoupled. The PES is the sum of the Coulomb-repulsion of the atom kernels and the Schrödinger equation of the electrons $\mathcal{H}\Psi = E\Psi$ [13,14]. The explicit calculation of the

energy E is not of interest, here. We assume the PES is given by a scalar function of the coordinates of the molecule.

Definition 1. Let \mathcal{K} be an open subset of \mathbb{R}^n . The function $E \in C^l(\mathcal{K}, \mathbb{R})$, $l \geq 2$ is an n -dimensional PES. \mathcal{K} is the configuration space of the PES. The derivative $G: \mathcal{K} \rightarrow \mathbb{R}^n$ with

$$G(\mathbf{x}) = dE(\mathbf{x}) = \left(\frac{\partial E}{\partial x_1}(\mathbf{x}), \dots, \frac{\partial E}{\partial x_n}(\mathbf{x}) \right)^T$$

is the gradient and the matrix $H(\mathbf{x}) \in \mathbb{R}^{n \times n}$ with

$$H(\mathbf{x}) = d^2E(\mathbf{x}) \left(\frac{\partial^2 E}{\partial x_i \partial x_j}(\mathbf{x}) \right)_{i,j=1}^n$$

is the Hessian with $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{K}$. The set $\mathcal{E}_\alpha = \{\mathbf{x} \in \mathcal{K} | E(\mathbf{x}) = \alpha\}$ is named equipotential surface.

The configuration space of a molecule is restricted. We assume at least a 2-fold differentiability of the PES for practical reasons – for the use of the diverse applications. The Hessian is symmetric and has the partition into eigenvalues and eigenvectors:

$$H = U \Lambda U^T, \quad (1)$$

with $U = (\mathbf{u}_1, \dots, \mathbf{u}_n)$ and $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n)$, with $H\mathbf{u}_i = \lambda_i \mathbf{u}_i$, $i = 1, \dots, n$.

Definition 2. A point $\mathbf{x} \in \mathcal{K}$ is nondegenerate if $\det H(\mathbf{x}) \neq 0$. In the contrary case it is degenerate. The index of a nondegenerate point $\mathbf{x} \in \mathcal{K}$ is the number of negative eigenvalues of $H(\mathbf{x})$. We write $\text{ind}(\mathbf{x})$, as well as $\text{ind}_2(\mathbf{x}) := \text{ind}(\mathbf{x}) \bmod 2$.

The value $\text{ind}_2(\mathbf{x})$ is equivalent to the sign of the determinant of the Hessian:

$$\det H(\mathbf{x}) > 0 \Leftrightarrow \text{ind}_2(\mathbf{x}) = 0 \quad \text{and} \quad \det H(\mathbf{x}) < 0 \Leftrightarrow \text{ind}_2(\mathbf{x}) = 1. \quad (2)$$

Definition 3. A point $\mathbf{x}_0 \in \mathcal{K}$ with $G(\mathbf{x}_0) = \mathbf{0}$ is named stationary point (StP). Ess \mathcal{K} is the set of all stationary points in \mathcal{K} . A nondegenerate stationary point, \mathbf{x}_0 , is:

- minimum, if $\text{ind}(\mathbf{x}_0) = 0$, or
- maximum, if $\text{ind}(\mathbf{x}_0) = n$, or
- saddle point of index i , if $\text{ind}(\mathbf{x}_0) = i$, $0 < i < n$.

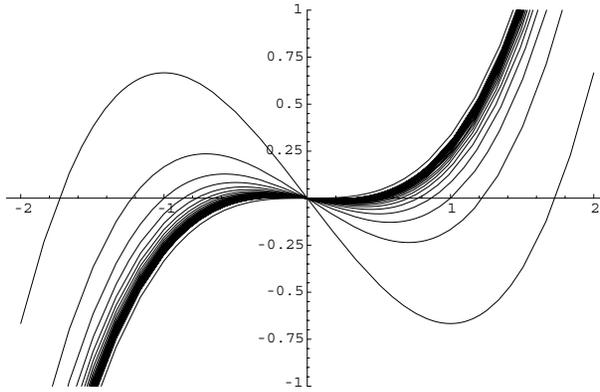


Figure 1. Dense subset of a function space.

We assume that no stationary point is degenerate, i.e. that for all $\mathbf{x} \in \mathcal{K}$ it holds the regularity condition

$$\|G(\mathbf{x})\| + |\det H(\mathbf{x})| > 0. \tag{3}$$

This convention will frequently occur in the sequel.

The subset \mathcal{F} of functions in $C^2(\mathbb{R}^n, \mathbb{R})$ fulfilling the condition is open and dense in the strong C^k -topology [15]. Dense means that for every function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ there is a series of functions f_n from \mathcal{F} with $\lim_{n \rightarrow \infty} f_n = f$. Figure 1 illustrates an example for $C^2(\mathbb{R}, \mathbb{R})$. For function $f(x) = \frac{1}{3}x^3$ the point $x = 0$ is a StP with $f'|_{x=0} = 0$, as well as it is degenerate because $f''|_{x=0} = 0$. For functions $f_n(x) = \frac{1}{3}x^3 - \frac{1}{n}x$ the StPs $x = \pm\sqrt{1/n}$ are nondegenerate with $f''_{n|x} = \pm\sqrt{1/n} = \pm 2\sqrt{1/n}$. It is $\lim_{n \rightarrow \infty} f_n = f$. Nondegenerate stationary points are isolated [16].

2.1. Valley-ridge-inflection points

A special subset of degenerate points can be interpreted to be the branching points of RPs, see figures 2 and 3.

Definition 4. A valley-ridge-inflection point (VRI) is located where the gradient is orthogonal to a zero eigenvector of the Hessian [17]. The subset of such points is $\text{Ext}(\mathcal{K})$.

The gradient does not lie in the kernel of the Hessian, and an augmented Hessian with gradient does not lift the defect of the rank:

$$\text{Ext}(\mathcal{K}) = \{\mathbf{x} \in \mathcal{K} | \text{rank}[H(\mathbf{x}), G(\mathbf{x})] < n\}. \tag{4}$$

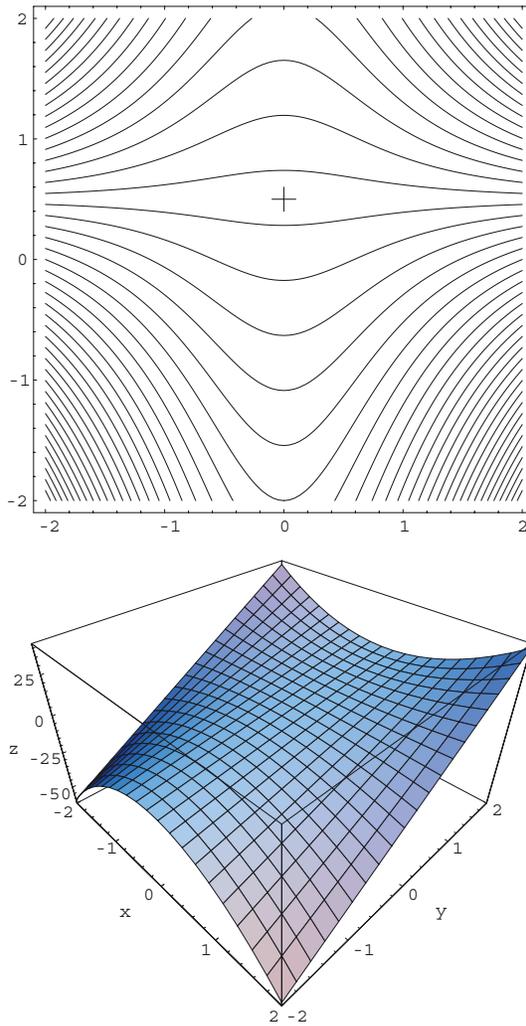


Figure 2. Symmetric valley-ridge-inflection point.

The bracket means matrix augmentation: $[H(\mathbf{x}), G(\mathbf{x})] \in \mathbb{R}^{n \times (n+1)}$. Note that not all VRI points are symmetric, see figure 3.

Definition 5. Points with $\det H(\mathbf{x}) = 0$ and $\text{rank}[H(\mathbf{x}), G(\mathbf{x})] = n$ are inflection points.

VRI points and inflection points are degenerate points of the PES. They are independent on any curve definition.

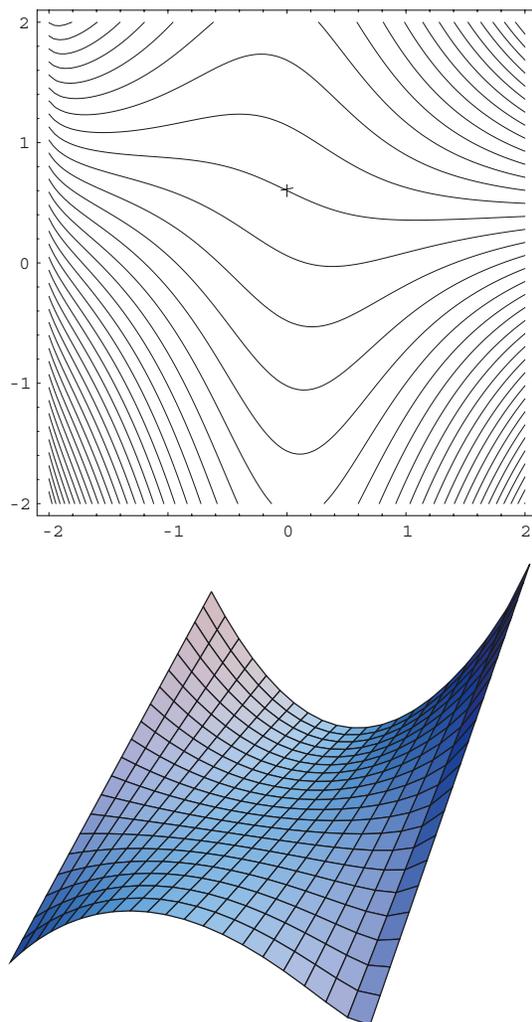


Figure 3. Un-symmetric valley-ridge-inflection point.

3. Steepest descent-IRC

The steepest descent (SD) from the SP in mass-weighted Cartesian coordinates is a simple definition of an RP, which is well-known as the intrinsic reaction coordinate (IRC) [3]. Using the arc-length s for the curve parameter, an SD curve $\mathbf{x}(s)$ is defined by the system of vector equations in n dimensions

$$\frac{d\mathbf{x}(s)}{ds} = -\frac{\mathbf{G}(\mathbf{x}(s))}{\|\mathbf{G}(\mathbf{x}(s))\|} = -\mathbf{w}(s), \quad \mathbf{x}(0) = \mathbf{x}_0, \quad (5)$$

where $\mathbf{G}(\mathbf{x})$ is the gradient vector of the PES and $\mathbf{w}(s)$ depicts the gradient direction with unit length. Starting at any \mathbf{x}_0 with $G(\mathbf{x}_0) \neq \mathbf{0}$, the solution of the differential equation leads monotonically decreasing to a minimum (or another deeper lying StP). The stationary points are the fixed points of the method. There the gradient is zero. In every non stationary point $\mathbf{x} \in \mathcal{K} \setminus \text{Ess}(\mathcal{K})$ the tangent of equation (5) is defined and there cannot be a branching [18].

The IRC is frequently used as a synonym for the MEP of the PES. But it has a serious imperfection: in one's imagination equation (5) permits to ascend from the minimum to the SP by changing $-\mathbf{w}(s)$ by $\mathbf{w}(s)$, however, it is impossible for practical use due to the funnel character of SD near the minimum [19] leading to the instability of an equation (5) "uphill" [9,20]. The IRC equation (5) explicitly needs the knowledge of the SP for the calculation of the steepest descent to the minimum.

SD curves can go down over a ridge, see figure 4. They do not always mirror the structure of valleys or ridges. The first derivatives of the PES are not sufficient to characterize the curvature of valleys, or ridges. In the un-symmetric case of the lower part of figure 4, a small variation of \mathbf{x}_0 with $G(\mathbf{x}_0) \neq \mathbf{0}$ for the SD in the upper valley ($x \approx 0$, $y \approx -1.5$) would not lead to the left side valley. Note that on a ridge the SD does not go along a minimum. However, it goes down along a maximum seen across the pathway: this contradicts the traditional definition of a minimum energy path [3].

We will extend the discussion. With Fukui [3] we start the IRC at an SP of index 1 in the direction of the decay vector, the eigenvector of the Hessian to the negative eigenvalue. We cannot locally prove that a point belongs to the IRC without passing the IRC from the beginning. If we follow Fukui, we can use the definition "Minimum Energy Path" if the tangential plane to the equipotential surface has an extremum on the IRC, namely a minimum [3].

The gradient is orthogonal to the equipotential surface $\mathcal{E}_{E(\mathbf{x})}$ in every point $\mathbf{x} \in \mathcal{K}$. At every point \mathbf{x} exists a tangential plane $T_x \mathcal{E}$ to equipotential surface $\mathcal{E}_{E(\mathbf{x})}$. All eigenvalues of the Hessian are positive in a neighborhood of the minimum. There the potential energy itself has a minimum on every tangential plane $T_x \mathcal{E}$ at \mathbf{x} , where $T_x \mathcal{E}$ touches $\mathcal{E}_{E(\mathbf{x})}$. At every point in the neighborhood of \mathbf{x} the equipotential surface $\mathcal{E}_{E(\mathbf{x})}$ lies below the tangential plane, because the equipotential surface is convex. Through every point goes a gradient descent. Thus, near the minimum every gradient descent is an MEP.

However, that does not hold for all points of the configuration space. If not all eigenvalues of the Hessian are positive, it need not be true. In figure 5 the 2D equipotential surface of a 3D PES is schematically shown. (The 3D PES is not illustrable! Examples of such 2D equipotential surfaces are explicitly shown in figure 4 of ref.[7].) The part in front of the picture is the region where all eigenvalues are still positive. It is assumed that at the point, where the gradient arrow is depicted, one eigenvalue of the Hessian becomes zero. There the equipotential

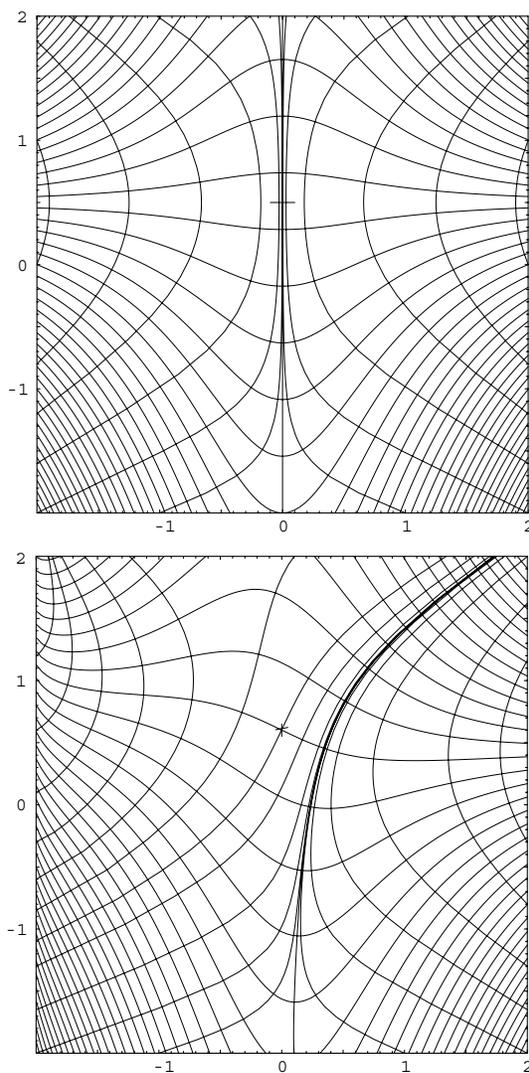


Figure 4. SDs near VRI points. The SDs start in regions with negative y values.

surface loses the convexity property if one goes along the IRC. The IRC itself loses the MEP property.

Going further, it is assumed throughout that one eigenvalue of the Hessian, being more or less orthogonal to IRC, is negative. In figure 6 the 2D equipotential surface of a 3D PES is schematically shown, in such a case. Now, the equipotential surface does not have a maximum at the crossing of the IRC and the tangential plane, but an SP. The IRC is not an MEP.

The number of eigenvalues being positive, negative or zero is equal to the number of coefficients of the quadratic form (1) $\mathbf{u}H\mathbf{u}$ being positive, negative or

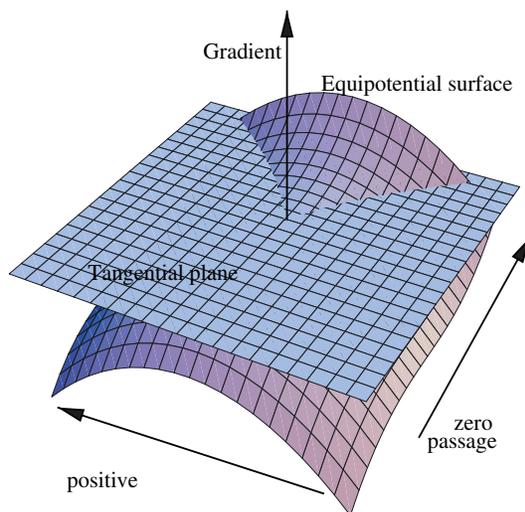


Figure 5. End of the MEP: tangential plane and 2D equipotential surface at a point where the equipotential surface crosses the convexity border.

zero. If the IRC leaves the valley region, one of the coefficients becomes zero, see figure 5, but the other coefficients stay positive. In figures 5 and 6 the arrows mark the corresponding subspaces. We conclude: an MEP can exist with the minimum property assumed by Fukui. Beginning at the SP of index 1, where all $(n - 1)$ eigenvalues orthogonal to the decay vector are positive, the IRC has the MEP property. However, the IRC can lose this property, cf [8].

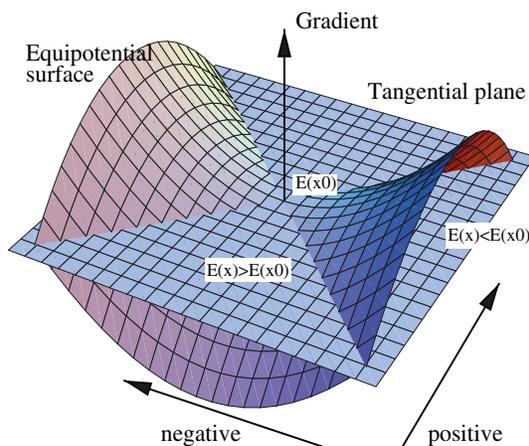


Figure 6. MEP? No, the tangential plane of a 2D equipotential surface (of a non shown 3D PES) is not minimum on the IRC.

4. Projection operator

It is $S^{n-1} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\}$ the unit sphere in \mathbb{R}^n . Elements from S^1 - the unit circle - are given as angles with point $(1, 0) \in S^1$ to be 0° .

We choose a column vector $\mathbf{r} \in S^{n-1}$ for the projection. It is a unit vector. Additionally, we use the transposed vector \mathbf{r}^T being a row vector. The dimension of \mathbf{r} is $(n \times 1)$ where that of \mathbf{r}^T is $(1 \times n)$. We form the dyadic product

$$\mathcal{D}_r = \mathbf{r} \cdot \mathbf{r}^T \quad (6)$$

which is an $(n \times n)$ matrix. \mathcal{D}_r projects with \mathbf{r} :

$$\mathcal{D}_r \mathbf{r} = (\mathbf{r} \cdot \mathbf{r}^T) \cdot \mathbf{r} = \mathbf{r} (\mathbf{r}^T \cdot \mathbf{r}) = \mathbf{r}, \quad (7)$$

where we use the unit length of \mathbf{r} in the scalar product. The projector which projects orthogonally to \mathbf{r} is

$$\mathcal{P}_r = \mathbf{I} - \mathcal{D}_r. \quad (8)$$

\mathbf{I} is the unit matrix. Application of \mathcal{P}_r to a vector, $\alpha \mathbf{r}$, parallel to \mathbf{r} gives

$$\mathcal{P}_r (\alpha \mathbf{r}) = \alpha (\mathbf{I} \cdot \mathbf{r} - \mathbf{r} (\mathbf{r}^T \cdot \mathbf{r})) = \alpha (\mathbf{r} - \mathbf{r}) = \mathbf{0}. \quad (9)$$

Vectors being orthogonal to \mathbf{r} are unchanged by \mathcal{P}_r .

5. RGF, Newton trajectories

It was proposed to choose a driving coordinate along the valley of the minimum, to go a step in this direction, and to perform an energy optimization of the residual coordinates [4]. Recently, the method was transformed into a new mathematical form to RGF [6, 7, 21]. The concept is that a selected gradient direction is fixed along the curve $\mathbf{x}(s)$

$$G(\mathbf{x}(s)) / \|G(\mathbf{x}(s))\| = \mathbf{r}, \quad (10)$$

where \mathbf{r} is the unit vector of the search direction. The search direction may correspond to an assumed start direction of a chemical reaction. Or, it may be the direction between the two minima of reactant and product, or the direction between the reactant and the assumed SP. Since \mathbf{r} is chosen to connect two specified minima the algorithm “knows” where it tries to go [5]. The property (10) is realizable by a projection of the gradient employing \mathcal{P}_r of (8). We pose

$$\mathcal{P}_r G(\mathbf{x}(s)) = \mathbf{0}. \quad (11)$$

\mathcal{P}_r is a constant matrix of rank $n-1$. It is $\text{Ker}(\mathcal{P}_r) = \text{lin}\{\mathbf{r}\}$ and $\text{Im}(\mathcal{P}_r) = \text{lin}\{\mathbf{r}\}^\perp$. \mathcal{P}_r projects a vector \mathbf{v} in direction of $\text{lin}\{\mathbf{r}\}$ and on $\text{lin}\{\mathbf{r}\}^\perp$.

Definition 6. The map $\mathcal{R}: \mathbb{R}^n \times S^{n-1} \rightarrow \mathbb{R}^{n-1}$, $\mathcal{R}(\mathbf{x}, \mathbf{r}) = \mathcal{P}_r G(\mathbf{x})$ will be called the reduced gradient, and $\mathbf{r} \in S^{n-1}$ will be called search direction. The equation

$$\mathcal{R}(\mathbf{x}, \bar{\mathbf{r}}) = \mathbf{0} \tag{12}$$

is for any fixed $\bar{\mathbf{r}} \in S^{n-1}$ the reduced gradient equation to the search direction $\bar{\mathbf{r}}$.

Based on the explicit definition, the predictor–corrector method of the reduced gradient following (RGF) [7] traces a curve [11] along its tangential vector by the derivative to obtain the tangent \mathbf{x}'

$$\mathbf{0} = \frac{d}{ds}[\mathcal{P}_r G(\mathbf{x}(s))] = \mathcal{P}_r \frac{dG(\mathbf{x}(s))}{ds} = \mathcal{P}_r \mathbf{H}(\mathbf{x}(s))\mathbf{x}'(s). \tag{13}$$

The RGF is a simple but effective procedure in order to determine all types of StPs [6]. In the general good-natured case, each RGF curve passes each StP. A whole family of RGF curves connects the extrema if we vary the search direction \mathbf{r} [9], see figure 7.

Definition 7. Be $r \in S^{n-1}$. We will name NT in \mathcal{K} to the direction \mathbf{r} the set:

$$\mathcal{T}_r(\mathcal{K}) := \{\mathbf{x} \in \mathcal{K} | G(\mathbf{x}) = r \|G(\mathbf{x})\|\}. \tag{14}$$

It is clear that \mathcal{T}_r is the set of solutions of equation (12). Or, in other words, it fulfills equation (10): the gradient points into the same direction. For every non

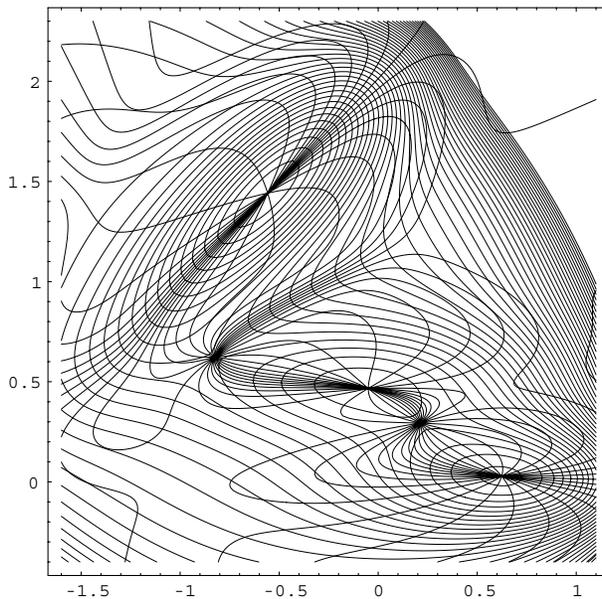


Figure 7. A family of Newton trajectories on Müller–Brown (MB) PES [10].

stationary point $\mathbf{x} \in \mathcal{K} \setminus \text{Ess}(\mathcal{K})$ the NT is given by the direction of the gradient. It means:

Proposition 1. [15] It holds

1. $\mathbf{x} \in \text{Ess}(\mathcal{K}) \Leftrightarrow$ for all $r \in S^{n-1}: \mathbf{x} \in \mathcal{T}_r(\mathcal{K}) \Leftrightarrow$ for $r, s \in S^{n-1}, r \neq s$: $\mathbf{x} \in \mathcal{T}_r$ and $\mathbf{x} \in \mathcal{T}_s$,
2. $\mathbf{x} \in \mathcal{K} \setminus \text{Ess}(\mathcal{K}) \Rightarrow \mathbf{x} \in \mathcal{T}_{G(\mathbf{x})/\|G(\mathbf{x})\|}$,
3. $\text{Ess}(\mathcal{K}) \neq \emptyset \Rightarrow \mathcal{T}_r(\mathcal{K}) \neq \emptyset$, for all $r \in S^{n-1}$.

On $\mathcal{K} \setminus \text{Ess}(\mathcal{K})$ a differentiable map $\underline{r}: \mathbb{R}^n \rightarrow S^{n-1}$, $\underline{r}(\mathbf{x}) := G(\mathbf{x})/\|G(\mathbf{x})\|$ exists, with $\mathcal{R}(\mathbf{x}, \underline{r}(\mathbf{x})) = \mathbf{0}$. The map is also named trajectory map [15].

Proposition 2. [15] Be $\bar{\mathbf{x}} \in \mathcal{K} \setminus \text{Ess}(\mathcal{K})$. Then is $\bar{\mathbf{x}} \in \text{Ext}(\mathcal{K}) \Leftrightarrow \text{rank } d\underline{r}(\bar{\mathbf{x}}) < n-1 \Leftrightarrow \dim \text{Ker } d\underline{r} > 1$. The kernel of $d\underline{r}(\bar{\mathbf{x}})$ is spanned by the tangent to the NT $\mathcal{T}_{\underline{r}(\bar{\mathbf{x}})}$ at $\bar{\mathbf{x}}$.

Figure 8 shows an NT for the PES of molecule HCN. The search direction is the angle coordinate of HCN. The curve in (x, y) -plane in figure 8 depicts the location of the H atom, the z -axis is the CN-distance of the configuration. The minimum HCN is at $(x, y) \approx (-1, 0)$, and the minimum HNC is at $(x, y) \approx (2, 0)$ [in Ångström].

5.1. Branches and components of Newton trajectories

NTs are smooth, parameterizable curves in non stationary points, where the rank of the matrix $\partial_{\mathbf{x}} \mathcal{R}(\mathbf{x}, r) = \mathcal{P}_r H(\mathbf{x})$ is $n - 1$ (i.e. maximal) [22]. It follows from the implicit function theorem. The condition is fulfilled if the rank of the augmented Hessian $[H, G]$ is maximal.

Definition 8. The set $\text{Ext}(\mathcal{K})$ of equation (4) is the set of extraneous singularities. All points in \mathcal{K} , which are neither in $\text{Ess}(\mathcal{K})$ nor in $\text{Ext}(\mathcal{K})$, are named regular.

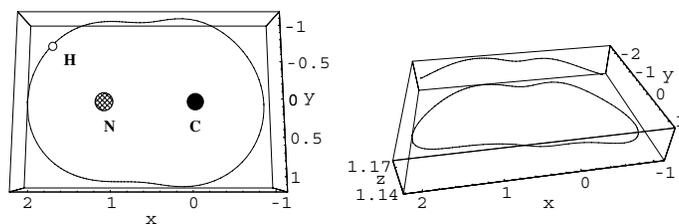


Figure 8. A Newton trajectory on the HF 6-31G*-PES of HCN.

The characterization of a nondegenerate point uses properties of the PES; but the characterization of a regular point uses properties of Newton trajectories. An inflection point (definition 5) is degenerate, however, in the general case, it is a regular point of the corresponding NT. If one takes off the singular points of an NT, one will get a set of smooth, parameterizable pieces of curves, see figure 9. Any such piece of a curve will be named a branch. The end points of open branches are the points of the boundary of \mathcal{K} or singular points. The NT of figure 9 is partitioned into three components: I and II are compact sets in the interior of \mathcal{K} , but III is open in \mathcal{K} . Component I contains two StPs, II does not contain any StP. Component III contains three StPs and one VRI point.

Proposition 3. [23] Every compact component of an NT, being disjunct to the boundary of \mathcal{K} , contains an even number or no stationary points.

Caution: The word “compact” in the hypothesis is crucial! Figure 10 shows the four NTs which cross the VRI points of the MB potential; the VRIs themselves are given in Table 1. The figure top right contains a compact component of an NT with two StPs, and the figure bottom right shows a compact component of an NT with four StPs.

5.2. Branin’s method

The reduced gradient approach shows an analogy to the mathematical theory of Branin [24], the global Newton method [15]. It utilizes the adjoint matrix A of the Hessian matrix H . This is defined as $((-1)^{i+j}m_{ij})^T$ where m_{ij} is the

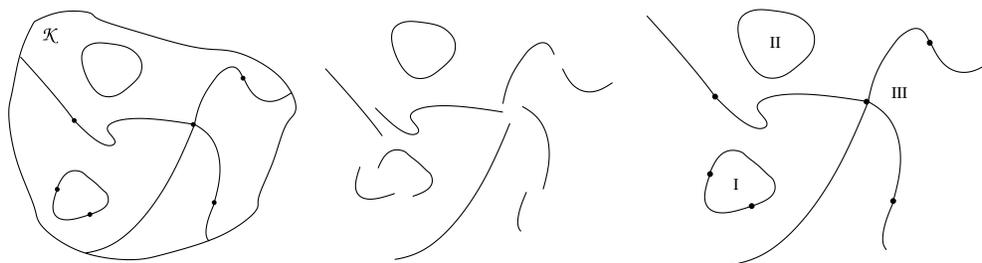


Figure 9. A Newton trajectory, its branches and components.

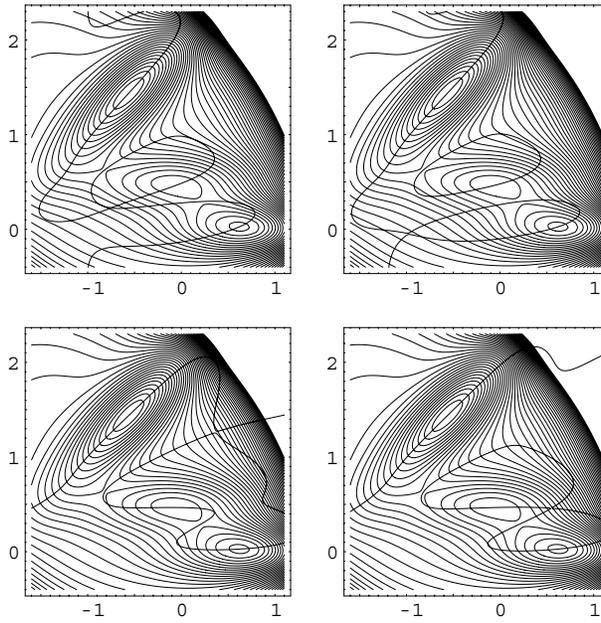


Figure 10. Singular NTs of the MB PES[10] which cross the VRIs. Corresponding search directions are 66.805°, 61.96°, 30.39°, 37.67° (from left above to right below).

minor of H obtained by deletion of the i th row and the j th column from H , and taking the determinant. The adjoint matrix satisfies the relation

$$H A = \det(H) I, \tag{15}$$

where $\det(H)$ is the determinant of H , and I is the unit matrix. The adjoint matrix A is used to define an autonomous system of differential equations for the curve $\mathbf{x}(s)$, where s is a curve parameter

$$\frac{d\mathbf{x}(s)}{ds} = + A(\mathbf{x}(s)) G(\mathbf{x}(s)). \tag{16}$$

Table 1
VRI points on MB-PES, and search direction of NTs.

| point | direction | x | y |
|------------------|-----------|----------|----------|
| VRI ₁ | 66.805° | -0.75002 | 0.22586 |
| VRI ₂ | 61.96° | -0.98072 | -0.04753 |
| VRI ₃ | 30.39° | 0.37250 | 1.26315 |
| VRI ₄ | 37.67° | 0.54859 | 0.45930 |

Proposition 4. Solutions of the Branin equation (16) are branches of Newton trajectories.

Proof. The phase portraits of solutions of the Newton method

$$\dot{\mathbf{x}} = -H(\mathbf{x})^{-1}G(\mathbf{x}) \tag{17}$$

and of equation (16) are equal, up to orientation [23]. Thus we can use equation (17) for the proof. Equation (12) is fulfilled for all \mathbf{x} with $\|G(\mathbf{x})\|_{\mathbf{r}} = G(\mathbf{x})$, i.e. for all points where the gradient is a pointer into direction \mathbf{r} . Along a trajectory of (17) holds

$$\frac{d}{dt}G(\mathbf{x}(t)) = H(\mathbf{x})\dot{\mathbf{x}} = -G(\mathbf{x}(t)).$$

The differential equation has the solution

$$G(\mathbf{x}(t)) = G(\mathbf{x}(0))e^{-t}.$$

It means that along a trajectory of equation (17), and equal to it along a trajectory of equation (16), the gradient has the direction $G(\mathbf{x}_0)$. Consequently, NTs are the same curves like RGF curves using the same gradient [7]. Because solutions of a differential equation are named trajectories, a solution of equation (16) will be called Newton trajectory (NT). \square

5.3. Stationary point: attractor or repeller?

In which direction the tangent of the Branin equation is directed in the neighborhood of an StP? We develop the gradient on the right hand side of equation (16) in a series at \mathbf{x}_0 with $G(\mathbf{x}_0) = 0$:

$$A(\mathbf{x})G(\mathbf{x}) \approx A(\mathbf{x}_0)H(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = \det H(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) . \tag{18}$$

If $\text{ind}_2 H(x_0) = 1$, thus if \mathbf{x}_0 is StP of odd index, then it holds in a neighborhood of \mathbf{x}_0 :

$$\dot{\mathbf{x}} \approx -|\det H(\mathbf{x}_0)|(\mathbf{x} - \mathbf{x}_0).$$

It means the tangent $\dot{\mathbf{x}}$ points to the stationary point; the point is an attractor. The “+” option in equation (16) is adapted for the search of StPs being attractors.

If $\text{ind}_2 H(x_0) = 0$, thus if \mathbf{x}_0 is StP of even index, then it holds in a neighborhood of \mathbf{x}_0 :

$$\dot{\mathbf{x}} \approx |\det H(\mathbf{x}_0)|(\mathbf{x} - \mathbf{x}_0).$$

It means the tangent $\dot{\mathbf{x}}$ points away; the point is a repeller. To search such an StP the “+” sign option of equation (16) has to be turned into a “-” option.

5.4. Turning point

Figure 11 compares the distinguished coordinate method (bottom) with an NT (top). If one starts at the global minimum A along the x direction, and minimizes orthogonally to x axis, then one follows the NT to 0° up to its first TP. There the distinguished coordinate method (bottom) leaps, but the NT (top) continues without any further ceremony. The x axis is the search direction, $r = (1, 0)$. We can follow the tangent without problems. The second panel is a copy

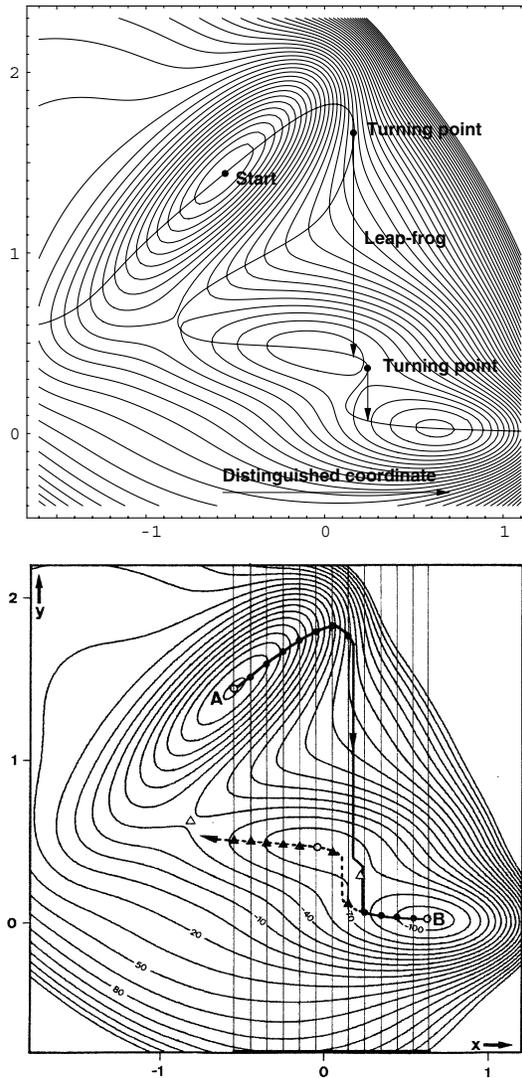


Figure 11. NT (top) and distinguished coordinate method [25], see text.

of figure 6 of ref. [25] (with permission). The pieces of the curves found there are from one and the same NT, see the part above.

Let $\bar{\mathbf{x}} \in \mathcal{T}_r$ be a regular point of an NT and $\mathbf{x}(t)$ is a parameterization of \mathcal{T}_r in the neighborhood of $\bar{\mathbf{x}}$. $\bar{\mathbf{x}} = \mathbf{x}(0)$ is the local maximum of the trajectory \mathcal{T}_r if $E(\bar{\mathbf{x}}) > E(\mathbf{x}(t))$, $\forall t \in (-\epsilon, \epsilon)$ and $t \neq 0$, i.e. $\mathbf{x}(t) \neq \bar{\mathbf{x}}$. The tangent τ to $\mathbf{x}(t)$ is at $\bar{\mathbf{x}}$ tangential to the equipotential surface $\mathcal{E}_{E(\bar{\mathbf{x}})}$ at $\bar{\mathbf{x}}$ and orthogonal to $G(\bar{\mathbf{x}})$, and thus, the tangent τ is orthogonal to the search direction r . For a local extremum holds that from $\tau(t)^T r > 0$ (or < 0) for $t \in (-\epsilon, 0)$ follows: $\tau(t)^T r < 0$ (or > 0) for $t \in (0, \epsilon)$.

Definition 9. A point is a TP of an NT (equation (16)) if the tangent $\dot{\mathbf{x}}(t)$ is orthogonal to the search direction r which is parallel to $G(\mathbf{x}(t))$.

For the NT of figure 11, \mathcal{T}_r , it is $\tau(t) = (0, -1)$ the tangent at the TPs.

6. Pseudo-convexity index

It is useful to have a criterion being simple and easy to calculate, if we follow curves through regions of valleys, or over ridges, on a way from minimum upon saddle. Such a criterion is the index of pseudo-convexity. It is defined by the well known Rayleigh quotient [26]. Usually, the quotient is employed for an eigenvalue approximation of a matrix M : $\varrho(v) = v^T M v / v^T v$. We use it with the adjoint matrix.

Definition 10. The pseudo-convexity index (pcx index) is the function

$$\xi: \mathcal{K} \setminus \text{Ess}(\mathcal{K}) \longrightarrow \mathbb{R}, \quad \xi(\mathbf{x}) = \frac{G(\mathbf{x})^T A(\mathbf{x}) G(\mathbf{x})}{G(\mathbf{x})^T G(\mathbf{x})}. \quad (19)$$

Figures 12 and 13 show examples for the pcx index over NTs. The NT unfolds on the x axis. The numbers are counted path-points. Places with “non continuous” behavior of the pcx index at StPs result from the singularity of the pcx index at StPs. Figure 12 shows some scalar values over the NT belonging to search direction 37.67° on the MB potential (see figure 10, the part bottom right). The NT leads through two minima, through two SPs, and through one VRI point. The message is: at TPs and VRIs the pcx index $\xi(\mathbf{x}(t))$ is zero. Figure 13 shows some scalar values on the HF 6-31G*-PES of HCN taken over the NT of Fig. 8. (Energy + 92.8) [in Hartree]: thin line; norm of gradient: dashed line; $0.03 \times$ pcx index: bold line. The path leads from HCN minimum over the transition structure to the HNC minimum. Neither a turning nor a bifurcation point emerge there. Thus, the pcx index is always positive.

At stationary points the pcx index is not defined. If $\mathcal{C}: \mathbb{R} \longrightarrow \mathbb{R}^n$ is a smooth curve and if $\mathcal{C}(0) = \mathbf{x}_0$ is a stationary point then the limit $\xi_0 := \lim_{t \rightarrow 0} \xi(\mathcal{C}(t))$ lies

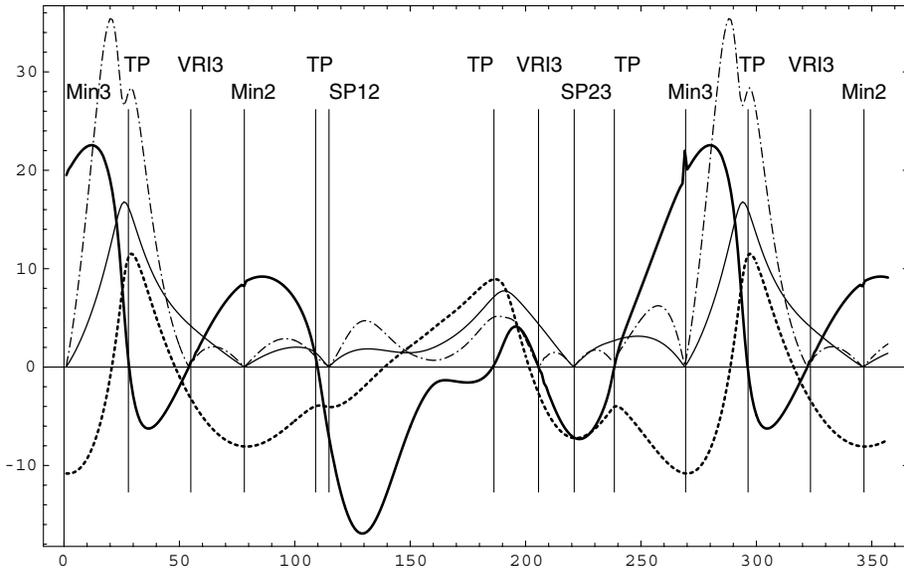


Figure 12. pex index (bold), $2 \times \|G(\mathbf{x}(t))\|$ (thin), $10 \times E(\mathbf{x}(t))$ (bold dashes), and $0.3 \times \|(AG)(\mathbf{x}(t))\|$ (thin dash-points) over an NT on MB.

between the maximal and minimal eigenvalue of $A(\mathbf{x}_0)$. Thus, for every minimum of the PES there is $\xi_0 > 0$, and for every maximum there is $\xi_0 < 0$, if $\text{ind}_2(\mathbf{x}_0) = 0$, correspondingly $\xi_0 > 0$, if $\text{ind}_2(\mathbf{x}_0) = 1$. Stationary points of index $1, \dots, n-1$ lie in the closure of the set

$$\Xi(\mathcal{K}) := \{\mathbf{x} \in \mathcal{K} \setminus \text{Ess}(\mathcal{K}) \mid \xi(\mathbf{x}) = 0\} . \tag{20}$$

There the sign of the limit ξ_0 depends on the direction of the incoming curve.

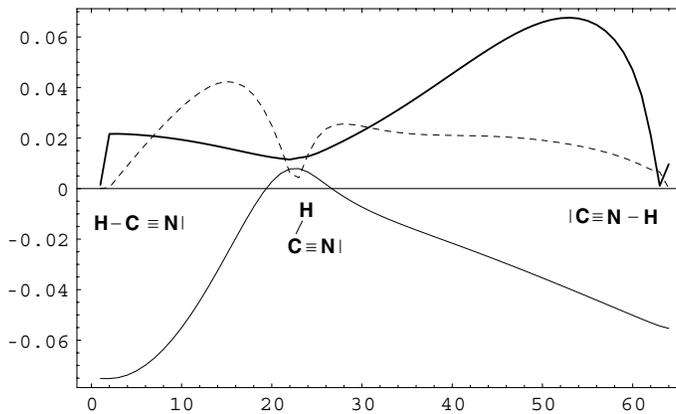


Figure 13. pex index (bold) over an NT as isomerization path on PES of HCN.

Table 2
Relation between index, eigenvalue λ ,
and pcx index ξ

| $\text{ind}_2(\mathbf{x})$ | λ | ξ |
|----------------------------|-----------|-------|
| 0 | + | + |
| 0 | - | - |
| 1 | + | - |
| 1 | - | + |

Definition 11. The set Ξ will be called boundary of pseudo-convexity.

Figure 14 shows the pseudo-convexity boundary Ξ on the MB potential. At the StPs of index one the boundary Ξ forms an “X” which was proposed for a leading line to find such SPs [1,27].

Proposition 5. On a gradient extremal (GE) [28,29] the relations of table 2 hold for the signs of the eigenvalue λ belonging to the gradient, for the index of the curve point \mathbf{x} , and for the pcx index ξ .

Proof. On a GE the gradient is eigenvector of H [29], but also of A : $AG = \mu G$ [7,30]. It follows from equation (19) that $\xi = \mu$ is the corresponding eigenvalue.

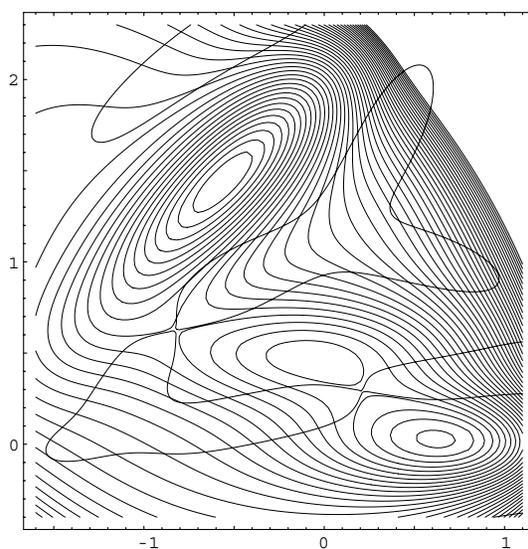


Figure 14. Pseudo-convexity boundary Ξ on MB potential.

But it holds for eigenvalues λ_i and μ_i of H , corresponding A : $\det H = \lambda_i \mu_i$ [7, 30]. Thus it holds on a GE that $\det H = \lambda \xi$. \square

Consequently, the pcx index is positive in the neighborhood of an SP of index one in direction of the negative eigenvalue, but the pcx index is negative in direction of the positive eigenvalues.

Conclusion 6. A gradient extremal crosses the boundary of pseudo-convexity Ξ in a VRI point.

Proof. On a GE holds $HG = \lambda_1 G$, $AG = \mu_1 G$, $\mu_1 G^T G = 0$ and thus $\mu_1 = \lambda_2, \dots, \lambda_n = 0$. It means that at least one eigenvalue of the $\lambda_2, \dots, \lambda_n$ is zero belonging to an eigenvector being orthogonal to the gradient. \square

It was proposed to calculate VRI points [8] using conclusion 6. The pcx index is, without the factor $1/\|G(\mathbf{x})\|^2$, the scalar product between the gradient in \mathbf{x} and the tangent to the NT through \mathbf{x} . The tangent is given by the Branin equation (16).

Proposition 7. The pseudo-convexity index ξ is zero at turning points and extraneous singularities of solutions of the Branin equation (16).

In other words: the set Ξ is the set of all prospective TPs of NTs. (Of course, not every NT has a TP).

Proof. There are two cases to treat because the pcx index is not defined in StPs.

1. $A(\mathbf{x})G(\mathbf{x}) = \mathbf{0}$. It means the gradient is eigenvector to eigenvalue zero of the adjoint Hessian. It follows from $\mu_i = \lambda_1 \cdots \lambda_{i-1} \cdot \lambda_{i+1} \cdots \lambda_n = 0$ that one of the eigenvalues of the Hessian orthogonal to the gradient is also zero, and thus, the rank of the augmented Hessian $[H(\mathbf{x}), G(\mathbf{x})]$ is not maximal. Consequently, \mathbf{x} is an extraneous singularity of the Branin method, a VRI point.
2. $G(\mathbf{x})$ is orthogonal to $A(\mathbf{x})G(\mathbf{x})$. We see from equation (16) that at \mathbf{x} the gradient is orthogonal to the tangent to the NT. It means that \mathbf{x} is a TP of the NT. \square

Now we need some definitions which connect the pcx index and convexity properties of the PES. They are prerequisites of further treatments.

Definition 12. [31] A set $\mathcal{K} \subset \mathbb{R}^n$ is named convex, if for all $x, y \in \mathcal{K}$ with $x \neq y$ the convex combination is lying in \mathcal{K} : $\lambda x + (1 - \lambda)y \in \mathcal{K}$ for $\lambda \in (0, 1)$.

Let \mathcal{K} be convex, $E: \mathcal{K} \rightarrow \mathbb{R}$. The set $\mathcal{L}_\alpha = \{\mathbf{x} \in \mathcal{K} \mid E(\mathbf{x}) \leq \alpha\}$ is named lower level set, cf. [32]. Let $\lambda \in (0, 1)$. A function E is named

- *convex* (cx) if $E(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda E(\mathbf{x}) + (1 - \lambda)E(\mathbf{y})$,
- *strictly convex* (s.cx) if $E(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda E(\mathbf{x}) + (1 - \lambda)E(\mathbf{y})$,
- *quasi-convex* (qcx) if \mathcal{L}_α is convex for all α ,
- *strictly quasi-convex* (s.qcx) if $x \neq y$:
 $E(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) < \max\{E(\mathbf{x}), E(\mathbf{y})\}$,
- *semistrictly quasi-convex* (ss.qcx) if $E(\mathbf{x}) \neq E(\mathbf{y})$:
 $E(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) < \max\{E(\mathbf{x}), E(\mathbf{y})\}$,
- *pseudo-convex* (pcx) if $E \in C^1(\mathcal{K}, \mathbb{R})$ and $(\mathbf{x} - \mathbf{y})^T \nabla E(\mathbf{y}) \geq 0$
 $\implies E(\mathbf{x}) \geq E(\mathbf{y})$,
- *strictly pseudo-convex* (s.pcx) if $E \in C^1(\mathcal{K}, \mathbb{R})$ and $(\mathbf{x} - \mathbf{y})^T \nabla E(\mathbf{y}) > 0$
 $\implies E(\mathbf{x}) > E(\mathbf{y})$.

Proposition 8. [31] The following relations hold for differentiable functions:

$$\begin{array}{ccccc}
 \text{s.cx} & \implies & \text{s.pcx} & \implies & \text{s.qcx} \\
 \downarrow & & \downarrow & & \downarrow \\
 \text{cx} & \implies & \text{pcx} & \implies & \text{ss.qcx} \implies \text{qcx}
 \end{array}$$

Definition 13. Two manifolds, M_1 and M_2 , are transversal in a common point \mathbf{x} , if the tangential spaces of the manifolds in \mathbf{x} span the whole \mathbb{R}^n , i.e., $\mathbb{R}^n = T_{\mathbf{x}}M_1 + T_{\mathbf{x}}M_2$.

M_1 and M_2 are named transversal, if they are transversal in every common point.

The definition 13 is illustrated in figure 15 in the 2D case. Let M be a compact, connected 1-codimensional differentiable manifold in \mathbb{R}^n , i.e., $\dim M = n - 1$. Then $\mathbb{R}^n \setminus M$ consists of two open components, one of which is bounded. It is the interior of M . We use for M an equipotential surface.

Definition 14. M is named global boundary to vector field \mathcal{N}_G of the tangents of Branin equation 16, if \mathcal{N}_G is transversal to M .

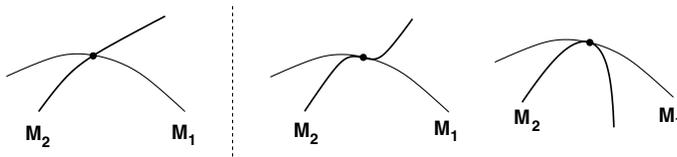


Figure 15. Transversal (left) and tangential (right) curves.

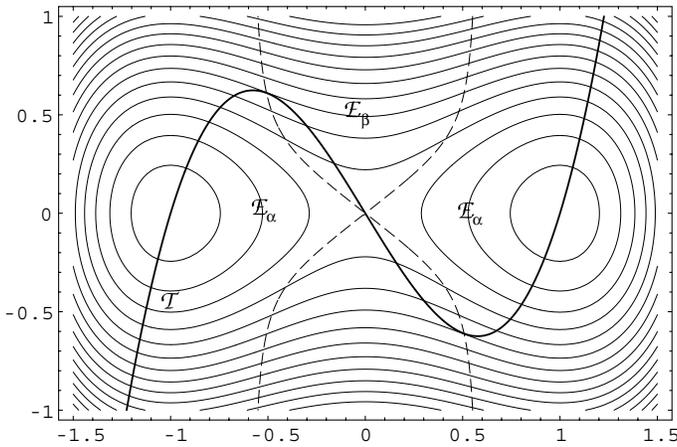


Figure 16. Global boundary \mathcal{E}_α to NT \mathcal{T} , see text.

In Figure 16 the NT \mathcal{T} (bold) transversally crosses the equipotential surface \mathcal{E}_α , and it tangentially touches \mathcal{E}_β . It holds $\xi(\mathbf{x}) = 0$ on the dashed line.

In every point $\mathbf{x} \in M$ the tangential space $T_x M$ spans the whole \mathbb{R}^n together with $\mathcal{N}_G(\mathbf{x})$. An equipotential surface is a global boundary of the vector field of the Branin equation (16), if there does not exist an NT which touches it tangentially. In figure 16 the two compact components of \mathcal{E}_α form a global boundary to every NT, however, not \mathcal{E}_β . The gradient is always orthogonal to the equipotential surface. Thus, an NT crosses the equipotential surface in a regular point $\bar{\mathbf{x}}$ transversally if it holds $\xi(\bar{\mathbf{x}}) \neq 0$, and tangentially if $\xi(\bar{\mathbf{x}}) = 0$.

Conclusion 9. A compact component of an equipotential surface \mathcal{E}_α is a global boundary of the vector field \mathcal{N}_G to the Branin equation (16) if the pcx index is not zero on \mathcal{E}_α .

Proposition 10. [15] Suppose that M is a global boundary for vector field \mathcal{N}_G . Then, the trajectory map $\underline{\Gamma}$ restricted to M is locally a diffeomorphism and maps M onto S^{n-1} .

If a compact component of an equipotential surface \mathcal{E}_α is a global boundary to \mathcal{N}_G , it is diffeomorph to the sphere in \mathbb{R}^n and, additionally, the gradient points on \mathcal{E}_α always in distinct directions. If $\mathbf{x}, \mathbf{y} \in \mathcal{E}_\alpha$ and $\frac{G(\mathbf{x})}{\|G(\mathbf{x})\|} = \frac{G(\mathbf{y})}{\|G(\mathbf{y})\|}$ then it holds $\mathbf{x} = \mathbf{y}$.

Proposition 11. Be M a 1-codimensional, compact manifold in \mathbb{R}^n , and be the field of normals on M diffeomorph to S^{n-1} , then the interior of M is convex.

Proof. We choose any direction \mathbf{v} and jam M with two $(n - 1)$ -dimensional hyperplanes orthogonal to \mathbf{v} like a vise. The hyperplanes are tangential planes to M . At the points of touch one normal vector directs to \mathbf{v} and the other directs to $-\mathbf{v}$. Every plane touches one point of M only, because otherwise, there would exist two equally directed normal vectors in different points of M , and the normal field would not be diffeomorph to S^{n-1} . If one moves one of the hyperplanes in direction \mathbf{v} into the interior of M , then the intersection is diffeomorph to S^{n-2} at the beginning. If, in a hypothetical case of the movement of the plane through the interior of M , the intersection does not become diffeomorph to S^{n-2} , then the plane would be in that point a tangential plane to M in a third point, but there the normal would be equally directed to \mathbf{v} or $-\mathbf{v}$. Consequently, the normal field to M would not be diffeomorph to S^{n-1} . Thus, every intersection of an $(n - 1)$ -dimensional plane having common points with the interior of M is diffeomorph to S^{n-2} and the interior of M is convex. \square

Conclusion 12. Let the pcx index not be zero on the boundary $\partial\mathcal{L}_\alpha = \mathcal{E}_\alpha$ of a compact component of a lower level set \mathcal{L}_α , and let the boundary not contain stationary points, then the component is convex.

Proof. The equipotential surface \mathcal{E}_α is a global boundary for \mathcal{N}_G because $\xi \neq 0$, and consequently, the gradient field, being for \mathcal{E}_α a field of normals, is diffeomorph to S^{n-1} on \mathcal{E}_α . With proposition 11 we get the conclusion. \square

Proposition 13 (Global boundary). [15] Let $E: \mathbb{R}^n \rightarrow \mathbb{R}$ and possible stationary points should not be degenerate. Additionally, let M be a global boundary to equation (16) and M does not contain extraneous singularities of E . Then holds:

- The interior of M contains no periodic trajectories of equation (16),
- M is diffeomorph to S^{n-1} ,
- The interior of M contains only one stationary point.

Figure 16 illustrates the proposition. Every component of the equipotential surface \mathcal{E}_α forms a global boundary to the field of tangent vectors to the NTs. Only on the line $\xi = 0$ the NTs have a TP. It means that compact components of equipotential surfaces which are not intersected by $\Xi = \{\mathbf{x} \in \mathcal{K} \mid \xi(\mathbf{x}) = 0\}$ form a global boundary to the tangent vector field belonging to equation (16). These components enclose only one minimum. However, the other equipotential surfaces (i.e. \mathcal{E}_β) enclose both minima, and they do not form a global boundary to \mathcal{N}_G .

Theorem 1. If the compact component of an equipotential surface \mathcal{E}_α does not contain stationary points, and if the pseudo-convexity index ξ is not zero on a corresponding lower level set $\mathcal{L}_\alpha \setminus \text{Ess}(\mathcal{K})$ without stationary points, then the PES, E , is strictly pseudo-convex (s.pcx) over this lower level set.

Proof. With the given assumptions it follows from conclusion 12 that every component of \mathcal{L}_γ , $\gamma < \alpha$ is convex, which is contained in the component \mathcal{L}_α . Let $\mathbf{x}, \mathbf{y} \in \mathcal{L}_\alpha$ and $(\mathbf{x} - \mathbf{y})^T G(\mathbf{y}) > 0$. Because $\mathcal{L}_{E(\mathbf{y})}$ is convex, it follows $E(\mathbf{x}) > E(\mathbf{y})$, and the theorem is shown. \square

The next step is to show that one can often continue the property of strictly pseudo-convexity uphill to an SP of index one. With that tool one can understand the character of the neighborhood of a curve, to be a valley or not. One can answer the question whether a curve leads through a valley or over a ridge employing the pcx index.

Proposition 14. Let $\mathcal{C}: (a, b) \rightarrow \mathbb{R}^n$, $\mathbf{x}(t) = \mathcal{C}(t)$ be a branch of an NT connecting a minimum, $\mathcal{C}(a)$, and a saddle, $\mathcal{C}(b)$, then, the following items are equivalent:

- \mathcal{C} is strictly pseudo-convex.
- The pcx index is larger than zero on \mathcal{C} .
- \mathcal{C} increases strongly monotonically.

Proof. The PES is given by $E(t) = E(\mathbf{x}(t))$ over the curve. It follows by the chain rule

$$\frac{\partial E}{\partial t} = \frac{\partial E}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}}{\partial t} = G(\mathbf{x}) \cdot \dot{\mathbf{x}} .$$

The minimum $\mathcal{C}(a) = \lim_{t \rightarrow a} \mathbf{x}(t)$ is a repeller of the vector field \mathcal{N}_G given by the Branin equation (16). The tangent $\dot{\mathbf{x}} = AG$ is a pointer away from minimum. Curve $\mathbf{x}(t)$ is passed in this direction. It holds for all $t \in (a, b)$

$$\frac{\partial E}{\partial t} = GAG = \xi \|G\|^2 > 0$$

and it follows from $t_2 > t_1$ that $E(t_2) > E(t_1)$. \square

Proposition 15. Let $U_{\bar{\mathbf{x}}}$ be an open neighborhood of a regular point $\bar{\mathbf{x}}$, and let $U_{\bar{\mathbf{x}}}$ not contain stationary points. The pcx index for all $\mathbf{x} \in U_{\bar{\mathbf{x}}}$ is not equal to zero. Then for every $\alpha \in \mathbb{R}$ the restriction of the trajectory map on the intersection of

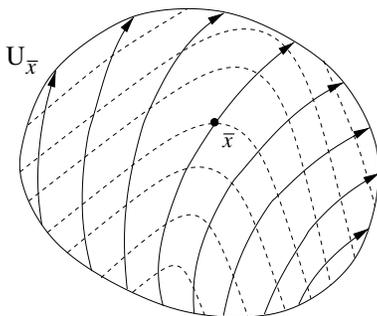


Figure 17. Convex neighbourhood of regular point \bar{x} with equipotential lines (dashed) and the field of NTs in a schematic view.

\mathcal{E}_α and $U_{\bar{x}}$, $\underline{r}: \mathcal{E}(\alpha) \cap U_{\bar{x}} \rightarrow S^{n-1}$ is a diffeomorphism on an open subset of S^{n-1} . (see figure 17.)

In other words: if the pcv index is not zero over an open subset of the configuration space, and if the subset does not contain stationary points, then every NT crosses every equipotential surface once only in the subset. It means that the gradient never has the same direction on the intersection, and there do not exist NTs to every direction in the subset. To prove the proposition 15 we use the proposition 2 to trajectory map.

Proof. We put $\mathcal{E}_\alpha(U_{\bar{x}}) = \mathcal{E}_\alpha \cap U_{\bar{x}}$ with $\alpha = E(\mathbf{x})$, $\mathbf{x} \in U_{\bar{x}}$. Because it holds $\xi(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in U_{\bar{x}}$, all points on $U_{\bar{x}}$ are regular. From the proposition 13 to global boundary follows that $\mathcal{E}_\alpha(U_{\bar{x}})$ is open for all α , because $U_{\bar{x}}$ should not contain stationary points. The vector field of NT \mathcal{N}_G crosses $\mathcal{E}_\alpha(U_{\bar{x}})$ transversally. The kernel of $d\underline{r}(\mathbf{x})$ is spanned by the tangent of the NT $\mathcal{T}_{\underline{r}(\mathbf{x})}$ through \mathbf{x} with proposition 2. For no $\mathbf{x} \in U_{\bar{x}}$ the kernel is contained in the tangential space $T_{\mathbf{x}}\mathcal{E}_\alpha$ to equipotential surface \mathcal{E}_α through \mathbf{x} . It follows that the rank of the Jacobi matrix is maximal, which is formed from the trajectory map $\underline{r}|_{\mathcal{E}_\alpha(U_{\bar{x}})}$ restricted to $\mathcal{E}_\alpha(U_{\bar{x}})$. It is $\text{rank } d\underline{r}|_{\mathcal{E}_\alpha(U_{\bar{x}})}(\mathbf{x}) = (n - 1)$ and $\dim \text{Ker } d\underline{r}|_{\mathcal{E}_\alpha(U_{\bar{x}})}(\mathbf{x}) = 0$. From the inverse function theorem follows that the restriction of \underline{r} onto $\mathcal{E}_\alpha(U_{\bar{x}})$ is a diffeomorphism and that $\underline{r}(\mathcal{E}_\alpha(U_{\bar{x}})) \subset S^{n-1}$ is open. \square

The open set $\mathcal{E}_\alpha(U_{\bar{x}}) = \mathcal{E}_\alpha \cap U_{\bar{x}}$ need not be connected. Proposition 15 is a supplement to proposition 10. The proof sounds analogous, see [15], and it uses that points with $\xi(\mathbf{x}) \neq 0$ are regular points of the PES.

Now, the next aim is to transfer the content of theorem 1 to the neighborhoods of proposition 15. It will be possible if the sets $\mathcal{E}_\alpha(U_{\bar{x}})$ can be seen to be a part of a compact set being diffeomorph to S^{n-1} and convex. Thus, if we can find a PES \tilde{E} which is equal with E on $U_{\bar{x}}$, and the levels $\mathcal{E}_\alpha(U_{\bar{x}})$ of it are subsets of equipotential surfaces fulfilling the assumptions of theorem 1. Consequently,

it is true if the $U_{\bar{x}}$ is near a minimum, and the equipotential surfaces are curved positively in all directions.

For $\text{ind}(\mathbf{x}) = 0$ the condition is fulfilled trivially.

For $\text{ind}(\mathbf{x}) = 1$ the number of negative eigenvalues of the adjoint matrix A is $n - 1$. From $\mathbf{u}_i^T H \mathbf{u}_i = \lambda \mathbf{u}_i^T \mathbf{u}_i < 0$ it follows $\mathbf{u}_i^T A \mathbf{u}_i = \mu \mathbf{u}_i^T \mathbf{u}_i > 0$, and vice versa. With Sylvester's law of inertia the number of positive and negative subspaces of a matrix is constant, respective of a linear operator, and it follows from $G^T A G = \xi \|G\|^2 > 0$, $G^T H G < 0$, and so $\mathbf{v}^T H \mathbf{v} > 0$ for all \mathbf{v} out of the tangential space to the equipotential surface staying orthogonally to G . So it holds:

Theorem 2. Be $U_{\bar{x}}$ an open and convex neighborhood of a regular point \bar{x} , which does not contain stationary points. Be the pcx index larger than zero for all $\mathbf{x} \in U_{\bar{x}}$, and $\text{ind}(\mathbf{x}) \leq 1$. Then the PES E is strictly pseudo-convex over $U_{\bar{x}}$.

6.1. Pseudo-convexity and structure of a valley

Around a minimum the pseudo-convexity boundary marks the end of pseudo-convexity, and in this sense the end of the valley character. The definitions of the pseudo-convexity index and the pseudo-convexity boundary are useful.

A sharpening of proposition 14 for a stricter convexity is not possible. A smooth curve connecting a minimum and an SP of index one always has at least one inflection point where the gradient has a local extremum. Such a curve cannot be convex. The pseudo-convexity makes that the curve increases strongly monotonically from minimum up to the SP. The change of the index at the inflection point is not important for answering the question of the character of the valley around the curve. An older, simpler formula, instead of equation (19) is given with H for A in the Rayleigh quotient [33]. Such a quotient would also include the "simple" inflection points. However, the pseudo-convexity index used here only shows the changes of the index taking place orthogonally to the gradient, thus the changes caused by the VRIs.

It is clear that the pseudo-convexity of a curve is not a sufficient criterion for a valley. Every steepest descent curve is pseudo-convex, however, it can go over a ridge. On the other hand, also the condition that the pcx index is larger than zero over any smooth curve is not a sufficient criterion for the pseudo-convexity of the curve. A smooth curve $C: (a, b) \rightarrow \mathbb{R}^n$, connecting a minimum, $C(a)$, and an SP of index one, $C(b)$, is strictly pseudo-convex if it increases strongly monotonically.

On the basis of the theorems 1 and 2 we find a general characterization for a valley with the help of the pseudo-convexity index.

Definition 15. Between a minimum, $C(a)$, and a saddle, $C(b)$, there is throughout a valley, if there is a smooth curve, a valley curve, connecting $C(a)$ and $C(b)$ and

fulfilling throughout $\xi > 0$, and for which the energy increases strongly monotonically.

From proposition 14 follows:

Theorem 3. The branch of an NT may connect a minimum and a SP of index one. It is a valley curve if the pcx index over the branch is larger than zero.

If one wants to know that an NT from minimum does not lead over a ridge, it is sufficient to know that its energy increases strongly monotonically. Complementary, an NT from minimum enters into a ridge region at its first TP. A branch being free of TPs of an NT always connects minimum and SP through a valley.

7. Discussion and conclusion

There is no method which always finds a “valley path” between minimum and SP. The reason is that such a valley does not always exist, see figure 8 of ref. [12], figure 4 of ref. [34] or figure 2 of ref. [35]. Additionally, a valley curve can exist but not an NT which is a valley curve, see figures 20 and 23 below. Our treatment is a hint that the definitions, which are given in this paper, are still too coarse. The mathematical definitions given here are general imaginations which one may have about valleys or ridges. The peculiarity of our ansatz is that its mathematical part is considered not as a game with deductive reasoning and symbols, but as a part of natural science, especially of chemistry, i.e. as an experimental treatment [36].

In this section we discuss the relation between the three curves of interest here, the IRC, the NT, and also the gradient extremal (GE), to describe a valley structure.

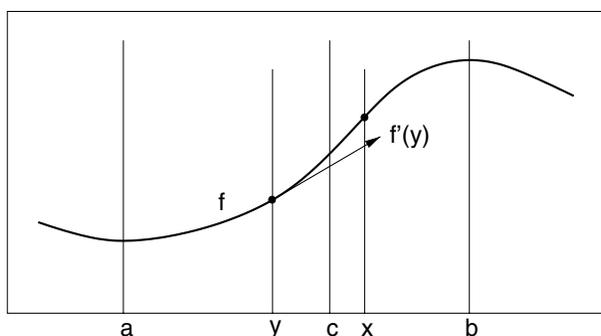


Figure 18. Pseudo-convex curve between a and b . f is convex between a and c , but not between c and b .

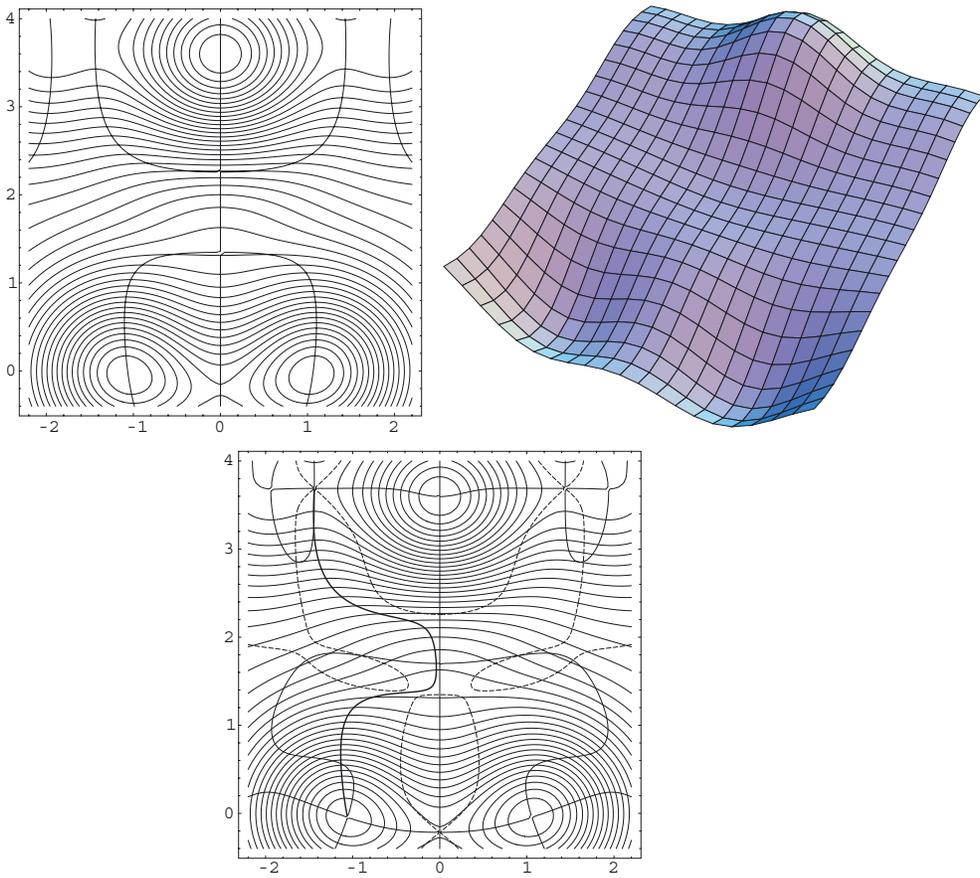


Figure 19. Poseidon PES [11], top: the NT belonging to 90° has two consecutive VRI points and forms a double trident. Bottom: an NT being valley trajectory (bold), pxc boundary (dashed) and GEs (small).

The lower part of figure 19 shows the “good” case for an NT being a valley trajectory from minimum to SP. Going through a valley region throughout means that along the increase of the RP we always have convex equipotential surfaces. NTs from the Branin method have special properties: a monotonically increasing NT automatically goes through a valley. Note that the GE does not connect the minimum and the SP of interest, like the given NT, in figure 19. At a turning point of an NT this NT has a local extremum of the energy, and it enters the complementary region. It changes from valley to ridge region, or vice versa. The property allows to construct a new index, the index of pseudo-convexity, using the Branin equation (16). The pxc index $\xi = G^T A G / (G^T G)$ answers the question for the valley- or ridge character of a region where the index of the Hessian is zero or one. For a Newton trajectory holds that if

it connects minimum and SP by a strongly monotonically increasing curve it fulfills $\xi > 0$. For other curves one may add the condition $\xi > 0$ to the condition of a strongly monotonically increasing curve to obtain a valley curve throughout.

Figure 20 shows the “bad” case, for one example, where no NT connects the minimum, being below in the center, and the SP in the right upper corner. However, there is a valley curve. From the minimum at $\approx (0.95, -1.05)$ leads a small valley region around $(1.45, -0.75)$ to the saddle of index one at $\approx (1.6, -0.3)$. We may draw a valley curve “by hand” through that region like in a children’s game. The scheme below shows that there is no branch of an NT from the minimum to the SP of index one in the right upper corner. The situation for the other higher SP of the example seems to be quantitatively similar, however, it is qualitatively different. There is also a small “corridor” with valley character, here around $(0.65, -0.85)$, being a bottle-neck, but the drawn NT exactly leads through the region of the corridor and connects minimum and the left upper SP with a valley trajectory.

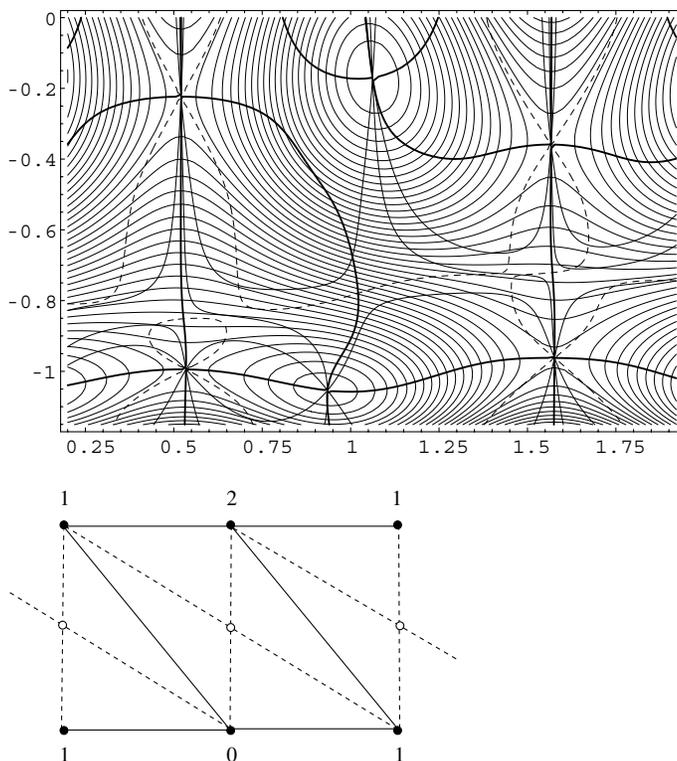


Figure 20. PES of Smeyer et al. [37], modified [11,34], top: GEs (bold), pcx boundary (dashed) and three NTs (small); bottom: scheme of singular NTs (dashed) with VRI points (empty). Given is also the index of the stationary points.

The question remains for conditions for the existence of an NT being a valley curve. Figure 21 schematically shows two pieces of an equipotential surface. We project the normalized tangents of NTs onto the tangential plane to the equipotential surface. We obtain vectors of length between 0 and 1. The projection has length 1 on the pcx boundary Ξ because there the tangent is in the tangential plane of the equipotential surface. The projection has length 0 if the tangent of an NT is orthogonal to the tangential plane of the equipotential surface, i.e., if it points with gradient direction. Then the equation (16) demands that the gradient is eigenvector of the adjoint matrix A , but of the Hessian H also. It is a GE shown on the left hand side. In the right part of figure 21 the projection of the normalized tangent always points in the same direction. One tangent points into the region with $\xi > 0$ but the other tangent points away from the region with $\xi > 0$. We meet the case in the right corridor of figure 20 at $(1.45, -0.75)$ as well. Thus, the right corridor need not be “permeable” to NTs being more or less orthogonal to the equipotential surfaces. However, in the other, left corridor of figure 20, there the pcx boundary Ξ is a global boundary of the vector field of the tangents and one can apply the proposition 13 to a global boundary. At a stationary point of the vector field we find a GE which crosses the equipotential surface.

One could speculate on the following

Conjecture 16. A valley curve meets the equipotential surfaces of the PES. For every equipotential surfaces it should be possible to find a global boundary inside the region of $\xi > 0$ of the vector field of the projected tangents. If every crossing point is inside the global boundary then a NT exits which is a valley curve.

Or similarly, if there is a valley curve such, that every equipotential surface which it crosses will also be crossed by a gradient extremal inside the same $\xi > 0$ region, then exits a NT which is a valley curve.

The conjuncture does not include that the GE itself connects minimum and SP, see figure 22, where in the left part the GE from SP is not a valley curve connecting the SP and the minimum. An NT is shown which is a valley curve. The right part of figure 19 also supports the conjuncture. There are very small



Figure 21. Projection of tangents of NTs on an equipotential surface which crosses two times the pcx boundary, Ξ . In the left case the projection has to be zero at a point GE, and the proposition 13 to global boundary is fulfilled. At GE the equipotential surface is crossed by a gradient extremal.

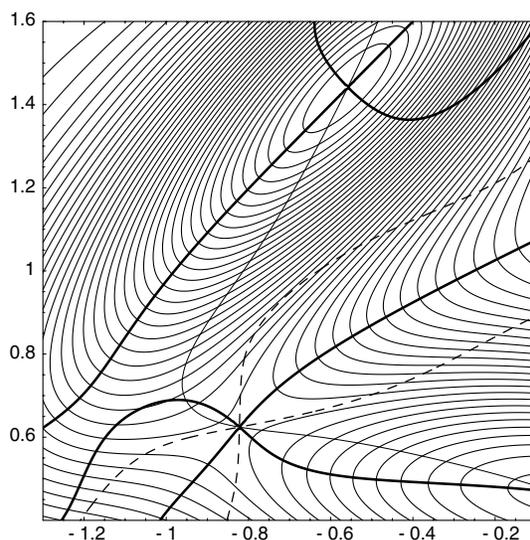


Figure 22. Gradient extremals (bold), one NT (thin) and pex boundary (dashed) on MB potential [10].

“corridors” for the valley region from minimum up to SP, but there are NTs finding their way through the corridors.

However, the hypothesis of the conjuncture 16 is not a sufficient prerequisite. This is demonstrated by a last figure 23. It is the slightly modified PES of Neria et al. [38]. Two minima at the sides are connected over an SP at $(0,0)$. There, the GE plays a lone hand being the valley curve throughout. None of the NTs connecting minimum and SP are valley curves.

The aim of our visualizations is to support analysis and interpretation of reaction pathways. Chemists have a long tradition in inventing and applying models for the analysis of molecular transformations taking place in a reaction. The development of new ideas, definitions, and methods for modeling an RP critically depends on visualization as an effective way to gain an understanding of a problem. For a long time the IRC was the model of choice of theoretical chemistry. It is simple, and it can also be understood to be a pre dynamical model of an RP [39,40]. However, the IRC from SP downhill can enter a ridge. Then it loses the property to be an MEP. One may exclude this situation by employing the pex index. If $\xi > 0$ on the whole IRC, it is a valley curve. The $\xi > 0$ condition gives one the possibility of a panoramic view over the RP of interest. The IRC is automatically decreasing throughout. If a NT monotonically increases between minimum and saddle then it is automatically a valley curve. The $\xi > 0$ condition is fulfilled. Thus, the NTs are well adapted to the problem treated in this paper: to enlighten the valley structure of the region around the reaction path.

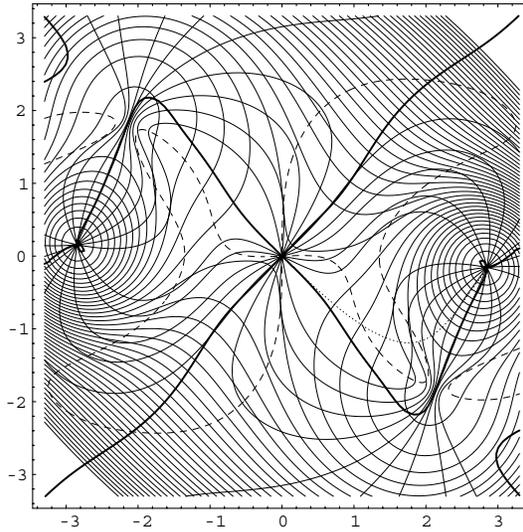


Figure 23. PES [38], modified [11]. Dashed line is the pcv boundary between valley and ridge. The IRC (thin points) crosses the ridge. A family of Newton trajectories is concentrated upon small gorges near $(\pm 2, \mp 2)$. However, none of them goes to the SP throughout in the valley. They cross the other ridge near the SP, or diverge. Only the gradient extremal (bold) is a valley curve throughout.

Appendix – List of model surfaces

Figures 2 and 3 use

$$E(x, y) = -5((y - 0.5)x^2 + y)$$

$$E(x, y) = -5((y - 0.5)x^2 + y + \ln(x + 2.1))$$

The Müller–Brown potential [10] is used in figures 7,10–12, 14, and 22. It is

$$E(x, y) = \sum_{i=1}^4 A_i \exp(a_i(x - x_i^0))^2 + b_i(x - x_i^0)(y - y_i^0) + c_i(y - y_i^0)^2$$

with $A = (-200, -100, -170, 15)$

$$a = (-1, -1, -6.5, 0.7)$$

$$b = (0, 0, 11, 0.6)$$

$$c = (-10, -10, -6.5, 0.7)$$

$$x^0 = (1, 0, -0.5, -1)$$

$$y^0 = (0, 0.5, 1.5, 1).$$

Figure 16 uses the potential $E(x, y) = 0.3(x + 1)^2(x - 1)^2 + y^2$.

Figure 19 is named “Poseidon” PES. It is

$$\begin{aligned}
 E(x, y) = & 0.02y^2 + 0.01x^2 - 0.2\delta\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}\right) \\
 & -\delta\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} 1.2 \\ 0 \end{pmatrix}\right) - \delta\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} -1.2 \\ 0 \end{pmatrix}\right) + \delta\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} 0 \\ 3.5 \end{pmatrix}\right) \\
 & -0.4\delta\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} 0 \\ -0.5 \end{pmatrix}\right) + 0.4\delta\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} -2 \\ 3.5 \end{pmatrix}\right) \\
 & +0.4\delta\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} 2 \\ 3.5 \end{pmatrix}\right)
 \end{aligned}$$

with the help of the dent function

$$\delta(\mathbf{x}, \mathbf{x}_0) := \exp(-(\mathbf{x} - \mathbf{x}_0)^2) \quad \text{and} \quad \mathbf{x} = (x, y), \quad \mathbf{x}_0 = (x_0, y_0)$$

Figure 20 is a PES of Smeyer et al. [37], modified by [34] and [11]. It is

$$\begin{aligned}
 E(x, y) = & 44730.4129 - 66786.5363 \cos(y) + 26352.6908 \cos(2y) \\
 & - 3117.3613 \cos(4y) + 659.3217 \cos(6y) - 111.5488 \cos(8y) \\
 & + 621.9640 \sin(3x) \sin(y) - 138.3050 \sin(3x) \sin(2y) - 500.0 \sin(3y) \\
 & - 7.7979 \sin(3x) \sin(6y) + 9.9258 \cos(6x) - 19.0681 \cos(6x) \cos(y) \\
 & + 41.8227 \sin(3x) \sin(4y) + 600 \cos(6x) \cos(2y) + 60x
 \end{aligned}$$

Figure 23 is a PES of Neria et al. [38], modified by [11]. It is

$$E(x, y) = \mathbf{c}(x^2 + y^2)^2 + xy - 9 \exp(-(x - 3)^2 - y^2) - 9 \exp(-(x + 3)^2 - y^2)$$

where we use $\mathbf{c} = 0.03$ and in [38] is used $\mathbf{c} = 0.06$.

Acknowledgments

The work was made possible through the financial support of the Deutsche Forschungsgemeinschaft. The authors thank Prof. D. Heidrich for stimulating discussions.

References

- [1] P.G. Mezey, *Potential Energy Hypersurfaces* (Elsevier, Amsterdam, 1987).
- [2] D. Heidrich, *The Reaction Path in Chemistry: Current Approaches and Perspectives* (Kluwer, Dordrecht, 1995).
- [3] K. Fukui, *J. Phys. Chem.* 74 (1970) 4161.
- [4] I.H. Williams and G.M. Maggiora, *J. Mol. Struct. (Theochem)* 89 (1982) 365.
- [5] R. Czerminski and R. Elber, *J. Chem. Phys.* 92 (1990) 5580.
- [6] W. Quapp, M. Hirsch, O. Imig and D. Heidrich, *J. Comput. Chem.* 19 (1998) 1087.

- [7] W. Quapp, M. Hirsch and D. Heidrich, *Theor. Chem. Acc.* 100 (1998) 285.
- [8] W. Quapp, M. Hirsch and D. Heidrich, *Theor. Chem. Acc.* 112 (2004) 40.
- [9] W. Quapp, *J. Comput. Chem.* 22 (2001) 537.
- [10] K. Müller and L.D. Brown, *Theor. Chim. Acta* 53 (1979) 75.
- [11] M. Hirsch, *Zum Reaktionswegcharakter von Newtontrajektorien, Dissertation*, (Fakultät für Chemie and Mineralogie, Universität Leipzig, (Dezember 2003). www.mathematic.uni-leipzig.de/MI/quapp/refs_mh
- [12] W. Quapp, *J. Theor. Comp. Chem.* 2 (2003) 385.
- [13] W. Kutzelnigg, *Einführung in die Theoretische Chemie* (VCH Verlagsgesellschaft, Weinheim, 1992).
- [14] W.J. Hehre, L. Radom, P.v.R. Schleyer and J.A. Pople, *Ab initio Molecular Orbital Theory* (Wiley, New York, Chichester, Brisbane, Toronto, Singapore, 1986).
- [15] H.T. Jongen, P. Jonker and F. Twilt, *Nonlinear Optimization in Finite Dimensions – Morse Theory, Chebyshev Approximation, Transversality, Flows, Parametric Aspects* (Kluwer, Dordrecht, 2000).
- [16] J. Milnor, *Morse Theory* (Princeton University Press, 5th Printing, 1973).
- [17] J. Baker and P.M.W. Gill, *J. Comput. Chem.* 9 (1988) 465.
- [18] W. Quapp, *J. Chem. Soc. Faraday Trans.* 90 (1994) 1607.
- [19] J.H. Hubbard and B.H. West, *Differential Equations, A Dynamical System Approach, Part I* (Springer, New York, 1991).
- [20] W. Quapp, *Chem. Phys. Lett.* 253 (1996) 286.
- [21] J.M. Anglada, E. Besalú, J.M. Bofill and R. Crehuet, *J. Comput. Chem.* 22 (2001) 387, J.M. Bofill and J.M. Anglada, *Theor. Chem. Acc.* 105 (2001) 463, R. Crehuet, J.M. Bofill and J.M. Anglada, *Theor. Chem. Acc.* 107 (2002) 130.
- [22] E.L. Allgower and K. Georg, *Numerical Continuation Methods – An Introduction* (Springer, Berlin, 1990).
- [23] I. Diener, *Globale Aspekte des kontinuierlichen Newtonverfahrens* (Habilitation, Göttingen, 1991).
- [24] F.H. Branin, *IBM J. Res. Dev.* 16 (1972) 504.
- [25] K. Müller, *Angew. Chem.* 92 (1980) 1.
- [26] A. Kielbasiński and H. Schwetlick, *Numerische Lineare Algebra* (Deutscher Verl. Wiss., Berlin, 1988).
- [27] P.G. Mezey, M.R. Peterson and I.G. Csizmadia, *Can. J. Chem.* 5 (1977) 2941.
- [28] S. Pancíř, *Collect. Czech. Chem. Comm.* 40 (1975) 1112.
- [29] M.V. Basilevsky and A.G. Shamov, *Chem. Phys.* 60 (1981) 347.
- [30] W. Quapp, *Optimization* 52 (2003) 317.
- [31] A. Cambini, E. Castagnoli, L. Martein, P. Mazzoleni and S. Schaible, *Generalized Convexity and Fractional Programming in Economic Applications*, (Springer, Berlin, 1990).
- [32] P.G. Mezey, *Chem. Phys. Lett.* 87 (1982) 277.
- [33] J. Kurchan and L. Laloux, *J. Phys. A* 29 (1996) 1929.
- [34] M. Hirsch and W. Quapp, *Chem. Phys. Lett.* 395 (2004) 150.
- [35] W. Quapp, *J. Mol. Struct.* 695–696 (2004) 95.
- [36] V.I. Arnold, *Arnold's Problems* (Springer, Berlin, 2004).
- [37] Y.G. Smeyer, M. Villa and M.L. Senent, *J. Mol. Spectrosc.* 177 (1996) 66.
- [38] E. Neria, S. Fischer and M. Karplus, *J. Chem. Phys.* 105 (1996) 1902.
- [39] J. González, X. Giménez and J.M. Bofill, *Phys. Chem. Chem. Phys.* 4 (2002) 2921.
- [40] J. González, X. Giménez and J.M. Bofill, *Theor. Chem. Acc.* 112 (2004) 75.