Reaction Pathways and Projection Operators: Application to String Methods

WOLFGANG QUAPP

Mathematical Institute, University of Leipzig, Augustus-Platz, D-04109 Leipzig, Germany

quapp@rz.uni-leipzig.de

Telephone: [49] 341-97 32153

Fax: [49] 341-97 32199

Web: www.mathematik.uni-leipzig.de/MI/quapp/string

Revised version of March 1, 2004, submitted to: Journal of Computational Chemistry

JCC-04-006

Keywords: Potential energy surface; reaction path; projected gradient; IRC; Newton trajectory; gradient extremal; chain methods

Subject area: Theoretical Methods and Algorithms

Proposed running head: Projection Operators & String Methods

Abstract

The reaction path is an important concept of theoretical chemistry. We discuss the definition with the help of diverse projection operators for the intrinsic reaction coordinate (IRC), for the following of the reduced gradient (RGF) or Newton trajectory (NT), and for the Gradient Extremal (GE). We describe different numerical schemes for the definitions in the context of string methods. It comes out that Newton trajectories are the best ansatz for a string method.

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

The starting point is a geometrically defined pathway which may serve as a reaction path. Geometrically defined means that only properties of the PES are taken into account, but 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 a reaction path ascends to the SP is an uncertainty of the RP definition. We use here the IRC,[3] the distinguished or driven coordinate method[4,5] in the modern form of RGF,[6,7] and the gradient extremal.[8] 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 the general case. (In a special theoretical case, the average of a set of molecular dynamics trajectories match very close the IRC.[9])

All the different forms of an MEP, above, can be defined with the help of projection operators. This tool is employed in string methods: the string is divided into a collection of points. They represent the MEP by a chain.[10,11] The proposal to applicate string methods to Newton trajectories (and also to the GE) is given here for the first time.

If the points of an initial chain evolve according to the potential forces of the chemical system then the problem emerges that we have to prevent all points from falling to the two minima. There are different possibilities to handle this: a “homogenization” of the chain by new points in the DDRP procedure,[12] or the inclusion of spring forces into the chain, for example in the nudged elastic band method (NEB).[13] A second problem is the affinity of gradient methods to zigzag. It is known that the pure string iteration for the IRC suffers from the problem in an enormous way. We find that the string iteration of a GE also shows the zigzagging in a mild form. The string iteration of a Newton trajectory does show the least zigzagging. It is locally defined by a very simple vector field, and the string method can be applied without spring forces or homogenization extras.

The described string procedure can be downloaded.[14] It is an adaption of a procedure of Stacho and Ban,[12] see also a flow-chart there.

2 Projection Operator

We choose a column vector $\mathbf{r}$ for the projection. It has to be a unit vector; in the contrary case it must be normalized. We additionally 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

$$\mathbf{D}_\mathbf{r} = \mathbf{r} \cdot \mathbf{r}^T \tag{1}$$

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

$$\mathbf{D}_\mathbf{r}\ \mathbf{r} = (\mathbf{r} \cdot \mathbf{r}^T) \cdot \mathbf{r} = \mathbf{r}\ (\mathbf{r}^T \cdot \mathbf{r}) = \mathbf{r} \tag{2}$$

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

$$\mathbf{P}_\mathbf{r} = \mathbf{I} - \mathbf{D}_\mathbf{r}\ . \tag{3}$$

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

$$\mathbf{P}_\mathbf{r}\ (\alpha\ \mathbf{r}) = \alpha\ (\mathbf{I} \cdot \mathbf{r} - \mathbf{r}\ (\mathbf{r}^T \cdot \mathbf{r}))\ = \alpha\ (\mathbf{r} - \mathbf{r})\ = \mathbf{0}\ . \tag{4}$$

Vectors being orthogonally to $\mathbf{r}$ are unchanged by $\mathbf{P}_\mathbf{r}$.

3 Steepest Descent – IRC

The steepest descent (SD) from the SP in mass-weighted Cartesian coordinates is a simple definition of a reaction path, which is well-known as the intrinsic reaction coordinate (IRC).[3]

Using the arc-length s for the curve parameter, a steepest descent curve $\mathbf{x}(s)$ is defined by the system of vector equations in n dimensions

$$\frac{d\mathbf{x}(s)}{ds} = - \frac{\mathbf{g}(\mathbf{x}(s))}{|\mathbf{g}(\mathbf{x}(s))|} = - \mathbf{w}(s) \tag{5}$$

where $\mathbf{g}(\mathbf{x})$ is the gradient vector of the PES and $\mathbf{w}(\mathrm{s})$ depicts the gradient direction with unit length. Numerically, the SD of the IRC is started by curve following at an SP of index one, a step in the direction of the decomposition vector. It is the eigenvector of the Hessian matrix with negative eigenvalue. The SD along the gradient, $-\mathbf{g}$, is calculated by discrete steps using the differential equation (5) with a step length factor.

The IRC is frequently used as a synonym for the MEP of the PES. But it has a serious imperfection: in one's imagination eq.(5) permits to ascent from the minimum to the SP by changing $-\mathbf{w}(s)$ by $\mathbf{w}(s)$, however it is not possible for practical use due to the funnel character of SD near the minimum[15] leading to the instability of an eq.(5) "uphill".[16,17] The IRC eq.(5) explicitely needs the knowledge of the SP for the calculation of the steepest descent to the minimum: start at the top and work your way down.

Because the SD follows the gradient of the PES, its tangent is parallel to the gradient and orthogonal to the equipotential hypersurface. Trivially, every SD fulfills the projector equation (4) where the tangent $\mathbf{t}(s)$ of the curve is used for the construction of the projector:

$$\mathbf{P}_{\mathbf{t}} = \mathbf{I} - \mathbf{t} \cdot \mathbf{t}^T \ . \tag{6}$$

If $\mathbf{x}(s)$ is the SD curve, then $\mathbf{t}(s) = \mathbf{w}(s)$ of eq.(5) is the tangent vector for the projector (6) and we have

$$(\mathbf{I} - \mathbf{w}(s) \cdot \mathbf{w}(s)^T)\ \mathbf{g}(\mathbf{x}(s)) = \boxed{\mathbf{P}_{\mathbf{w}(s)}\ \mathbf{g}(\mathbf{x}(s)) = \mathbf{0}\ .} \tag{7}$$

Additionally to eq.(5), we have a second definition of the SD by the projector equation (7). There is also an old second method to find the IRC by this ansatz. Liotard and Penot,[10] and Ulitzky and Elber[11] have chosen a smooth curve which connects two minima. If it is outside the IRC then the eq.(7) is not fulfilled at least in some points. Because, there is one and only one curve which goes through the SP and is SD. We take a resulting vector $\mathbf{p}$=$\mathbf{g}^{\perp}$ after the projection with the tangent of the curve

$$\mathbf{P}_{\mathbf{t}(s)}\ (-\mathbf{g}(\mathbf{x}(s))) = \mathbf{p}(s)\ . \tag{8}$$

The orthogonal part to $\mathbf{t}$ of $\mathbf{-g}$ (downhill) is used for a numeric approximation of a better curve. The chain of points is varied along the $\mathbf{p}$ vectors under some iterations, and there is hope for convergence at the IRC. Usually, if successful, the iteration loops find the IRC without pre-knowledge of the SP, as it is needed in the SD method above for (5). The way to locally follow the IRC from a minimum uphill using eq.(7) is not possible, too. If we search for the tangent of the curve by derivative to the curve parameter of the projector equation $\mathbf{P_w}\,\mathbf{g} = \mathbf{0}$, we do not get an equation for the tangent of the curve. By contrast, only a zero operator results throughout, see the Appendix. Thus, the eqs.(7) and (8) need a global string which connects the basins of two minima.

Fig. 1 shows an initial chain of points and the approximated IRC between the left minimum and the SP for the demonstration of this method. The surface is a simplified 2D model of the PES of malonaldehyde.[7,18] It represents the H transfer in the O-H-O fragment. If d_1 is the O_1-H distance, and d_2 is the O_2-H distance, the coordinates used are $x = d_1 - d_2$ and y is motion along the O_1-O_2 stretch. The PES is

$$E(x, y) = 2\ y + y^2 + (y + 0.4\ x^2)\ x^2\ . \tag{9}$$

The use of the model PES is a reduced model of the full molecular PES. The dimension of 21 internal degrees of freedom is shorten to two coordinates only. All couplings between these two coordinates and the other coordinates are neglected. In malonaldehyde this is questionable because the OH stretches have strong mode mixing.[18]

Fig. 1 shows the result of iterations of variations along $\mathbf{p}$ vectors to the given initial arc of points. Note that the start chain was chosen not fully symmetrically between the minima. The slight asymmetric points of the initial guess are given in Table I. The left result is a fine chain for the IRC from SP. However, the right hand side also shows a typical situation: the chain zigzags. (For the demonstration, the vector field $\mathbf{p}$ is iteratively used without homogenization,[12] without spring forces,[13] or without other reparameterization schemes.[19]) On the right string, the projectors totally puzzle the initial chain. The calculation is made by the step length factor 0.125 for the $\mathbf{p}$'s, and at the given picture the calculation stops by a convergence criterion for the highest point of the chain, the SP, after $k = 13$ loops. If one uses another convergence criterion, for example the difference between the points of loop $k-1$ and loop k, then the zigzagging may fall into a quasi stable "trap" after a longer run of loops, or it can diverge at all. Note that the "good" case on the left hand side, and the "bad" case on the right side of the chain are obtained by one and the same calculation.

4 RGF curves

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.[6,7,20] The concept is that a selected gradient direction is fixed along the curve $\mathbf{x}(s)$

$$\mathbf{g}(\mathbf{x}(s))/|\mathbf{g}(\mathbf{x}(s))| = \mathbf{r}, \tag{10}$$

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 $\mathbf{P_r}$ of (3). We pose

$$\boxed{\mathbf{P_r}\,\mathbf{g}(\mathbf{x}(s)) = \mathbf{0}\ .} \tag{11}$$

$\mathbf{P_r}$ is a constant $n{\times}n$ matrix of rank $n{-}1$. Based on the explicit definition, the predictor-corrector method of RGF[7] follows a curve (11) along its tangential vector by the derivative to obtain the tangent $\mathbf{x}'$

$$\mathbf{0} \ = \ \frac{d}{d\,s}[\mathbf{P_r}\ \mathbf{g}(\mathbf{x}(s))] = \mathbf{P_r}\ \frac{d\,\mathbf{g}(\mathbf{x}(s))}{d\,s} = \mathbf{P_r}\ \mathbf{H}(\mathbf{x}(s))\ \mathbf{x}'(s)\ . \tag{12}$$

The matrix $\mathbf{H}$ is the Hessian. The RGF is a simple but effective procedure in order to determine all types of stationary points.[6] In the general good-natured case, each RGF curve passes each stationary point. RGF curves are defined by a constant gradient, and a full family of them connects the extrema if we vary the search direction $\mathbf{r}$.[17]

The projector (11) can be used in a string method at every initial chain point, without any further derivative. If it is not zero, choose the downhill direction along

$$\mathbf{p} = -\mathbf{P_r}\ \mathbf{g} \ = -(\mathbf{I} - \mathbf{r}\cdot\mathbf{r}^T)\ \mathbf{g}\ . \tag{13}$$

Fig. 2 shows the vector field on PES (9) of the projector to the direction E_y=0, thus $\mathbf{r}$=$(1,0)^T$. From a mathematical point of view, the vector field looks somehow boring. The projector simply becomes

$$\mathbf{P_r} = (\delta_{ij} - r_i r_j)_{i,j=1,2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} . \tag{14}$$

The search direction $\mathbf{r}$ is that direction which connects the two minima, and the corresponding zero line of the field is a good approximation of a reaction path. It is clear that the iterated application of the projector leads to "straight forward" moves of the points of any initial string onto the searched trajectory, if the step length is chosen moderately enough. The resulting chain converges to the RGF to direction $\mathbf{r}$ with an accuracy as exact as one poses it (in the 2D case). For Fig.2, we start with the linear straight chain of points between the minima, and a step length factor of 0.25. We need 6 iteration loops for an accuracy for the sum of the positive differences between all the points of the 5th and the 6th chain of 0.0125.

RGF as a method to move forward in the suspected reaction valley is tested to be an effective tool in determining the next SP,[17] if starting at a minimum. Here we find that its use in a string method is very effective, too. Note that the application of the string method for projector (13) only needs the gradient, but not the Hessian matrix of the PES.

Fig. 3 shows RGFs starting at the left minimum to three different search directions for $\mathbf{r}$. The middle curve (large dashes) is again the curve to direction E_y=0. The full curve is the RGF to search direction $E_x+2\,E_y$=0. It is also a monotonous increasing curve between the minima and the SP, however, its suitability for a reaction path is, in the left part, already a borderline case – if we have in mind also a dynamical behavior. The third curve is the RGF to search direction E_x-0.05 E_y=0. (Curve with small dashes: the direction is much steeper than the valley directions of the corresponding minima.) The left branch is a curve with a turning point (TP) for the energy, which turns from ascent to descent, and there is a whole part of points where the energy is higher than the energy of the SP. This curve is not a model for a minimum energy reaction path. The turning point case[4,17,21] divides RGFs into those which can serve as loose RPs, and others: if the RGF does not contain a TP at the pathway from minimum up to the SP, it may be used as an RP model.[22] (However, in any case, an RGF curve may serve for a pre-dynamical trajectory of a reaction.[23])

Example of the Müller-Brown PES

We use the MB potential,[17,21] see **Fig. 4**, for a second test of the string method for RGFs. We start with a straight line between the two outer minima, M_1 and M_3, giving the search direction $\mathbf{r}$. We use the damping factor of 0.125, and a convergence criterion of 0.04 for the convergence of two loops. An initial chain of 33 points is used to get a continuous approximation of the path through 3 minima and 2 SPs, and the string method for the RGF to the direction between the two minima is used. Projector (13) results in the pathway of **Fig. 5**, a way over the known SPs and through the intermediate minimum, M_2. On the right hand side, the RGF between the stationary points is near to the IRC, however, the left part of the RGF curve between the deepest minimum, M_1, and the SP_1 is a different path, in comparison to the IRC. But it can also be seen as a model of an RP. The convergence to level 0.04 is slowly but straight forward. It needs $k = 9$ loops.

A stronger criterion makes more loops without a qualitative better result at SP_1. However, a weaker criterion results in a stop of the calculation where the chain is still not near enough at the SP_1. A larger step length factor causes zigzagging near the minimum M_3. The convergence to the RGF curve is very quickly between the minima M_2 and M_3. But the convergence is slowly on the part between minimum M_1 and SP_1. Because, that part in the main valley of the minimum M_1 goes along a very oblique direction in comparison to the search direction $\mathbf{r}$. Calculating that part alone tolerates a very larger step length factor: using only the first 8 points of the initial chain (seen from M_1) and a damping factor of 0.4 then the string method converges in 5 loops with a criterion of 0.025. At the other side, using only the last 14 points of the initial chain and a factor of 0.08, then the method converges nearly perfect with 4 loops, to a criterion of 0.02, against the true RGF curve between the minima M_2 and M_3.

Higher dimensional example

We use the Lennard-Jones (LJ) cluster[24] of seven Argon atoms, LJ_7,[25–27] for a 21-dimensional example of the string method. There are two minima, a pentagonal bipyramid at -16.51 energy units, and a capped octahedron at -15.94 energy units, see Table II. The search direction $\mathbf{r}$ is the direction between the minima. We start with a straight line guess between the minima: the 3D coordinates (x, y, z) of atoms 1-7 of the two minima are used in a linear interpolation. We use the small damping factor of 0.03 corresponding to the PES of the floppy cluster and the higher dimension of the example. The convergence criterion is 0.015 for the SP convergence. A chain of 12 points is used, and the string method for that RGF with projector (13) gets the MEP over the known SP of -15.44 energy units in $k = 10$ loops, see Fig. 6. The geometry of the highest point of the resulting chain is given for the SP in Table II. The convergence to level 0.015 is still straight forward: note that the string method for RGFs is simply an SD in an $(n-1)$-dimensional subspace. This method also suffers from a mild zigzagging.

The PES of the LJ cluster is used in the full Cartesian coordinates including the possibility of overall translation and rotation of the cluster. It does not mean any difficulties for the method, because the downhill steps of the projected gradient do use only non-zero parts of the gradient.

5 Gradient Extremal

Pancíř[28] and Basilevsky/Shamov[29] formulated local criteria for describing a valley floor line. Pancíř determined two conditions which he assumed to be obviously given:

(i) The energy must increase along all directions perpendicularly to the direction of the valley floor line.

(ii) The curvature of the energy surface along the direction of the valley must be less than the curvature along any other direction.

If the norm of the gradient forms a minimum along points of an equi-potential surface, E($\mathbf{x}$)=const., i.e. along all directions perpendicular to the gradient,[8,29] a point of gentlest ascent of a valley is found. The measure for the ascent of the function $E(\mathbf{x})$ is the norm of the gradient vector, the functional

$$\sigma(\mathbf{x}) = \frac{1}{2}|\mathbf{g}(\mathbf{x})|^2 \ . \tag{15}$$

Using the normalized gradient $\mathbf{w}(s)$ of eq.(5) and defining the projector

$$\mathbf{P}_{\mathbf{w}(s)} = \mathbf{I} - \mathbf{w}(s)\ \mathbf{w}(s)^T \ , \tag{16}$$

the requirement for an extremal value of σ is expressed by[38,39]

$$\mathbf{P}_{\mathbf{w}(s)}\nabla\ \sigma(\mathbf{x}(s)) \ = \ \boxed{\mathbf{P}_{\mathbf{w}(s)}\mathbf{H}(\mathbf{x}(s))\ \mathbf{g}(\mathbf{x}(s)) \ = \ 0\ .} \tag{17}$$

Setting $\lambda = \mathbf{w}^{\mathbf{T}}\,\mathbf{H}\,\mathbf{w}$, it results in the known eigenvector relation

$$\mathbf{H}(\mathbf{x})\ \mathbf{g}(\mathbf{x}) = \lambda(\mathbf{x})\mathbf{g}(\mathbf{x})\ . \tag{18}$$

Curves $\mathbf{x}(s)$ defined by (17) or (18) consisting of such points on consecutive equi-hyper-surfaces are termed *gradient extremals*.[8] On a GE, eq.(17) reduces to the trivial $\mathbf{P}_{\mathbf{w}}\ \mathbf{g} = \mathbf{0}$. The GE forms an isolated curve in the configuration space. It does not form fields of curves as do the SD lines.

In **Fig. 7** we show the GEs (short dashes) for the function of example (9). The valley-floor GE to the smallest eigenvalue follows the streambed of the surface, where the GE to the second eigenvalue follows a ridge or a cirque. The vector field to eq.(17) is[38]

$$\mathbf{p} = -\mathbf{P}_{\mathbf{w}}\ (\mathbf{H}\ \mathbf{g})\ . \tag{19}$$

It is also shown. If one uses it for the variation of a chain in a string method, there are two possibilities: the sign of the vector may cause that it points to the GE: then the method may converge. Or the vector goes off the curve: then the method diverges away from the GE. The application of a string method is more tricky for projector (19). But it seems possible, at least on some parts of the GE.

If starting on PES (9) with the straight line connection between the minima the convergence happens only in the region of the interval $x \in (-1.3,\ 1.3)$ around the SP. Also the ridge being below the SP is enfolded by the approximation. Points from the region of the minima are cleared away by the approximation. The vector field $\mathbf{p}$ shifts the end points of the initial chain. The reason is the projector itself which projects orthogonally to the gradient, thus parallel to the equipotential lines. Points besides the valley floor, at the slope, are not moved downhill. They are moved quasi parallel to the reaction path.

In the case of Fig. 7 we start with the result of the string method for the RGF of Section 4 given in Fig. 2. Also with such a good initial guess, the resulting chain is collected around the SP. The convergence is slow and the iteration cannot be driven under every pre-given level of accuracy. The iteration zigzags near the given picture shown in Fig. 7. The step length factor used is 0.125, and the iteration is stopped after k=18 loops with a convergence criterion of 0.05 for the difference of two loops.

An ansatz of a string method with the long known projector (19) adapts only the valley of the SP region. It is not able to adapt the streambed of the valley ground GE near the minima. But there is a dual GE projector which may be used alternately, see Section 8.

6 Branin's method

The RGF approach shows an analogy to the mathematical theory of Branin,[30] the global Newton method.[31] It utilizes the adjoint matrix $\mathbf{A}$ of the Hessian matrix $\mathbf{H}$. This is defined as $((-1)^{i+j} m_{i\,j})^T$ where $m_{i\,j}$ is the minor of $\mathbf{H}$ obtained by deletion of the i^{th} row and the j^{th} column from $\mathbf{H}$, and taking the determinant. The adjoint matrix satisfies the relation

$$\mathbf{H}\,\mathbf{A} = Det(\mathbf{H})\,\mathbf{I}\ , \tag{20}$$

where $Det(\mathbf{H})$ is the determinant of $\mathbf{H}$, and $\mathbf{I}$ is the unit matrix. The adjoint matrix $\mathbf{A}$ is used to define an autonomous system of differential equations, similar to (5), for the curve $\mathbf{x}(s)$, where s is a curve parameter

$$\frac{d\mathbf{x}}{ds} = \pm\,\mathbf{A}(\mathbf{x})\ \mathbf{g}(\mathbf{x}). \tag{21}$$

The "+" option is used for searching stationary points with an odd index (SPs with an odd number of negative eigenvalues of the Hessian), where the "−" option searches for stationary points with an even index. Because solutions of a differential equation are named trajectories, the solutions of (21) are named Newton trajectories (NT). They are the same curves like RGF curves using the same gradient.[7]

7 The TASC method

The TASC method[32] replaces the constant search direction $\mathbf{r}$ in eq.(11) of the predictor-corrector method of RGF by a direction which is changed during the iteration process: the tangent direction of the previous curve point iteratively becomes the search direction for the next point of the curve. The procedure is named the tangent search concept (TASC). Curiously, the projector with the unit vector $\mathbf{t}$=$\mathbf{x}'(s)/|\mathbf{x}'(s)|$, from the eq.(12), is formally the same projector as the IRC projector in eq.(6)! All calculations of the predictor-corrector method of TASC were done by eqs. (11) and (12) with a fixed $\mathbf{t}$. In the tangent derivative for (12) one approximates a "constant" $\mathbf{P_t}$ matrix in the current step. Predictor and corrector will work locally to search for an RGF curve to the gradient direction $\mathbf{t}(s)$=$\mathbf{r}$. This is an approximation, but it works and really goes along the streambed line. Usually, TASC changes the corresponding NT after the predictor step. This results in a self-consistency on the valley floor GE. Details are given in Refs.[32–34] There is compelling proof that the method converges to the GE, if appropriate conditions are fulfilled.[34] Additionally to the GE projector (17) we obtain from the proof of the convergence of TASC[34] that points on a GE fulfill a further projector equation. Using the normalized Newton tangent of eq.(21) of a NT

$$\mathbf{a}(s) = \frac{\mathbf{A}(\mathbf{x}(s))\mathbf{g}(\mathbf{x}(s))}{|\mathbf{A}(\mathbf{x}(s))\mathbf{g}(\mathbf{x}(s))|} \tag{22}$$

and defining the projector

$$\mathbf{P}_{\mathbf{a}(s)} = \mathbf{I} - \mathbf{a}(s)\ \mathbf{a}(s)^T\ , \tag{23}$$

the GE fulfills the alternate projection equation[34]

$$\boxed{\mathbf{P}_{\mathbf{a}(s)}\mathbf{g}(\mathbf{x}(s))\ =\ 0\ .} \tag{24}$$

The reason is that the matrices $\mathbf{H}$ and $\mathbf{A}$ have the same eigenvectors. If $\mathbf{g}$ is an eigenvector of $\mathbf{H}$ then it is also an eigenvector of $\mathbf{A}$ and eq.(24) reduces to $\mathbf{P_w}\ \mathbf{g} = \mathbf{0}$ again.

8 An alternate GE string method

Fig. 8 shows the vector field of projector (24) on the PES (9). The zero lines are, analogous to Fig. 7, the GEs. However, outside the GEs the fields are different. So, the vector fields $\mathbf{P_a}\ \mathbf{g}$ and $\mathbf{P_w}\ \mathbf{H}\,\mathbf{g}$ can be understood to be dual tasks for a GE calculation. In Fig. 8 it is demonstrated that the projector $\mathbf{P_a}$ is better adapted to a string method. The valley GE leading through the minima, and over the SP, is throughout an attractor for $-\mathbf{P_a}\ \mathbf{g}$, however, the cirque GEs from the minima, as well as the ridge GEs from the SP are repellors of the field. An enfolding of "false" GEs as in Section 5 is probably not, using the dual projector. But the repellor property of the ridge GE over the SP may cause a small drawback of this projector version: chain points may be moved away from the SP. The chain can be diluted at the SP.

With the same damping factor of 0.125 of Section 5 we obtain the valley floor GE in Fig. 8 after k=9 loops. The convergence is straight forward. The convergence criterion used is the very fine value of 0.005 for the difference of the two last chains. Note that a damping factor for $\mathbf{p}$ is always needed. The pure gradient projection may be too large for a useful numeric variation of the chain. It is a question of experience to select the best damping factor.

A change of the sign used in the projectors $\mathbf{P_a}$ and $\mathbf{P_w}$ changes the index of the GEs which we may search for. In Fig. 7 the plus sign can be used for the calculation of the cirque GE through the minima, if a good initial chain is taken before. In Fig. 8 the plus sign can be used for calculation of the ridge GE only, also after taking a good initial chain.

A combination of the two projectors $\mathbf{P_a}\,\mathbf{g}$ and $\mathbf{P_w}\,\mathbf{H}\,\mathbf{g}$ in a twofold calculation can be the following: Determine by $\mathbf{P_a}\,\mathbf{g}$ the global valley floor over the SP, and refine the SP region by $\mathbf{P_w}\,\mathbf{H}\,\mathbf{g}$.

9 Discussion

The different types of algorithms for locating SPs are based on four methods: RP methods, grid search, conjugate gradients, and quasi-Newton techniques.[35] We use the first method: we search reaction paths. Per definition, an RP leads over the SP, and we obtain automatically the SP if we are able to calculate the whole RP by a chain of points. In this paper we try to construct the RP by a string method consisting in to define a guess curve divided in segments. We have given a general, unified view to IRC, NT, and GE by projector operators which refine the guess chain into the RP.

Note that no dynamical behavior of the molecule is used. However, using the RP Hamiltonian method for IRC[37] or NTs,[23] it is possible to study nuclear dynamics with the "sole" restriction to move the system along the RP. Normally, with this type of constraint, one recovers many important molecular dynamic effects. The reformulation of the RP Hamiltonian has been demonstrated the advantage to evaluate first an RP and second to compute a constrained dynamics on this path.[23]

The calculation of the IRC can only go on if one uses a globally given string. Because, the trivial equation $\mathbf{P_w}\,\mathbf{g} = \mathbf{0}$ holds in every point of the configuration space. The equation alone cannot be a help for the finding of the pathway across the SP. If combined with the tangent of any curve, the projector can serve as a search tool of the IRC. However, in the numerical realization, the procedure suffers from strong zigzagging. This is due to the instability of the "direction that is adjusted on the fly", see also the Appendix.

– One way out is the DDRP method.[12] It uses for the string variation the simpler "projector" **P=I** applied to **-g**. It needs a "homogenization" of the chain. It has a high grade of reliability but also has a higher amount of calculation. A plus of the method is that it moves (or fixes) the start and the end point of the string into the two minima of reactant and product.

– A second remedy is a higher order organization of the string smoothing by an essentially nonoszillatory scheme.[19] It also works, but again, it has a higher amount of organization of the chain points on the string.

– The NEB method[13] also stabilizes the zigzagging by an higher amount of the use of spring forces. Their weighting and other problems emerge,[36] see also ref.[27] and further references therein.

If one simplifies the "variable" tangent projector of the IRC search to a constant projector defined by a constant search direction $\mathbf{r}$, for example by the (normalized) direction between the two minima, there results a projector which leads to a Newton trajectory, locally point by point. Every point of an initial string over the PES is moved locally and independently from the other points of the string, to its final place on the Newton trajectory. Like the IRC, most of the Newton trajectories can serve as model of a reaction path.[22] Thus, the calculation of Newton trajectories may be a serious alternate to the IRC using the string method. The definition of the projector is simpler than that of the IRC, the convergence of the string method is quite quicker than that of the IRC and with quite less zigzagging, and the saddle point is located on all Newton trajectories, with more safety than the SP on the special IRC line, in comparison to all other steepest descent lines.[17] The "global" action character of the projector $\mathbf{P_r}$ allows a piece-wise use of the strings at any part of the trajectory: a finer chain can be used to localize any part, for example the SP, with a smaller point distance. The application of the string method for NTs needs only the calculation of the gradient, point by point, but not the Hessian matrix.

The string treatment can also be tested for projectors of the Gradient Extremal, however, the vector fields of those projectors usually do not lead to a successful variation of an initial string to the "next" GE. Usually, some obstacles remain. There are regions and GEs to special indices where the method works well, without the tremendous third derivatives of the previous GE method.[38,39] But there are also regions where the variation of the initial chain clears away the points from the next GE. Additionally disturbing is the crossing of n GEs at the SP, if n is the dimension of the PES, which usually does not allow that the resulting string connects two minima. It is possible that the resulting "string" of an iteration of the GE also folds around a ridge GE which belongs to the SP. The use of the dual projector (24) can avoid the problem. At all, GEs are calculable by the string method, but its use needs a higher amount of organization and scrupulousness. The shift of the sign of the two projectors, and their dual use allow the calculation of individual pieces of GEs. Because of a mild zigzagging, often only coarse approximations are possible. To employ the projector (24) one needs the Hessian matrix "only", and its transcription to the adjoint matrix, however, one does not need third derivatives to calculate a GE.[38]

Appendix

We show the impossibility to use the gradient projector uphill for the tangent in a predictor step. With the normalized gradient of eq.(5) and the projector of eq.(16) we start with the trivial relation which holds for all curves of steepest descent

$$\mathbf{P}_{\mathbf{w}(s)}\ \mathbf{g}(\mathbf{x}(s)) = \mathbf{0}\ . \tag{25}$$

Because, the tangent to the steepest descent is the gradient itself. If one formally searches for the tangent to such a curve by the derivative to the curve parameter, one gets

$$\frac{d\ \mathbf{P}_{\mathbf{w}(s)}\ \mathbf{g}(\mathbf{x}(s))}{ds} = \sum_{i=1}^{n} \frac{\partial\ \mathbf{P}_{\mathbf{w}(s)}\ \mathbf{g}(\mathbf{x}(s))}{\partial x_i}\ \frac{d\ x_i}{ds}\ , \tag{26}$$

but every partial derivative of the sum is zero

$$\frac{\partial \, \mathbf{P}_{\mathbf{w}(s)} \, \mathbf{g}(\mathbf{x}(s))}{\partial x_i} \; = \; \mathbf{P}_{\mathbf{w}} \, \frac{\partial \, \mathbf{g}}{\partial x_i} \; - \frac{\partial \, \mathbf{w}}{\partial x_i} \, \mathbf{w}^T \, \mathbf{g} - \; \mathbf{w} \, \frac{\partial \, \mathbf{w}^T}{\partial x_i} \, \mathbf{g} \; = \; \mathbf{0} \tag{27}$$

using the relation[38]

$$\mathbf{P}_{\mathbf{w}} \, d\,\mathbf{g} = |\mathbf{g}| \; d\,\mathbf{w} \; , \tag{28}$$

as well as

$$\mathbf{P}_{\mathbf{w}}{}^T = \mathbf{P}_{\mathbf{w}} \quad \text{and} \qquad d\,(\mathbf{g}^T) \; \mathbf{P}_{\mathbf{w}} \; = |\mathbf{g}| \; d\,(\mathbf{w}^T) \; . \tag{29}$$

Thus, the derivative to the curve parameter s in eq.(25) becomes singular at every regular point of the configuration space, and a predictor step along the tangent becomes undefined. Every direction of a "tangent"-predictor goes to a next point which fulfills the eq.(25). The IRC cannot be defined locally. An equation like (12) does not exist for it.

Acknowledgments

The work was made possible through the financial support of the Deutsche Forschungsgemeinschaft. The author thanks M. Hirsch for stimulating discussions and the referees for suggestions and nice comments.

References

1. Mezey, P. G. Potential Energy Hypersurfaces; Elsevier: Amsterdam, 1987.

2. Heidrich, D. (ed.) The Reaction Path in Chemistry: Current Approaches and Perspectives, Kluwer: Dordrecht, 1995.

3. Fukui, K. J Phys Chem 1970, 74, 4161-4163.

4. Williams, I. H.; Maggiora, G. M. J Mol Struct (Theochem) 1982, 89, 365-378.

5. Czerminski, R.; Elber R. J Chem Phys 1990, 92, 5580-5601.

6. Quapp, W.; Hirsch, M.; Imig, O.; Heidrich, D. J Comput Chem 1998, 19, 1087-1100.

7. Quapp, W.; Hirsch, M.; Heidrich, D. Theoret Chem Acc 1998, 100, 285-299.

8. Hoffman, D. K.; Nord, R. S.; Ruedenberg, K. Theoret Chim Acta 1986, 69, 265-280.

9. González J.; Giménez, X.; Bofill, J. M. Phys Chem Chem Phys 2002, 4, 2921-2926.

10. Liotard, D. A.; Penot, J. P. In: DellaDora, J.; Demongeot, J.; Lacolle, B. (eds.) Study of Critical Phenomena, Springer: Berlin, 1981; Page 213-221.

11. Ulitzky, A,; Elber, R. J Chem Phys 1990, 92, 1510-1511.

12. Stacho, L. L.; Ban, M. I. Computer Chem 1993, 17, 21-25.

13. Henkelman, G.; Jónsson, H. J Chem Phys 2000, 113, 9978-87.

14. Quapp, W. E-mail: quapp@rz.uni-leipzig.de
    Web: www.mathematik.uni-leipzig.de/MI/quapp/string

15. Hubbard, J. H.; West, B. H. Differential Equations, A Dynamical System Approach, Part I; Springer: New York, 1991.

16. Quapp, W. Chem Phys Lett 1996, 253, 286-292.

17. Quapp, W. J Comput Chem 2001, 22, 537-540.

18. Yagi, K.; Taketsugu, T.; Hirao, K. J Chem Phys 2001, 115, 10647-10655.

19. Ren, W. Comm Math Sci 2003, 1, 377-384;
E, W.; Ren, W.; Vanden-Eijnden, E. Phys Rev B 2002, 66, 052301-1-4.

20. Anglada, J. M.; Besalú, E.; Bofill, J. M.; Crehuet, R. J Comput Chem 2001, 22, 387-406;
Bofill, J. M.; Anglada, J. M. Theor Chem Acc 2001, 105, 463-472;
Crehuet, R.; Bofill, J. M.; Anglada, J. M. Theor Chem Acc 2002, 107, 130-139.

21. Müller, K.; Brown, L.D. Theor Chim Acta 1979, 53, 75-93.

22. Quapp, W. J Theor Comput Chem 2003, 2, 385-417.

23. González, J.; Giménez, X.; Bofill, J. M. J Chem Phys 2002, 116, 8713-8722.

24. Lennard-Jones, J. E. Proc R Soc London, Ser A 1924, 106, 463-477.

25. Matro, A.; Freeman, D. L.; Doll, J. D. J Chem Phys 1994, 101, 10458-10463.

26. Chaudhury, P.; Bhattacharyya, S. P.; Quapp, W. Chem Phys 2000, 253, 295-303.

27. Trygubenko, S. A.; Wales, D. J. J Chem Phys 2004, 120, 2082-2094.

28. Pancíř, S. Collect Czech Chem Comm 1975, 40, 1112-1118.

29. Basilevsky, M. V.; Shamov, A.G. Chem Phys 1981, 60, 347-358.

30. Branin, F. H. IBM J Res Develop 1972, 16, 504-522.

31. Jongen, H. T.; Jonker, P.; Twilt, F. Nonlinear Optimization in Finite Dimensions – Morse Theory, Chebychev Approximation, Transversality, Flows, Parametric Aspects; Kluwer: Dordrecht, 2000.

32. Quapp, W.; Hirsch, M.; Heidrich, D. Theoret Chem Acc 2000, 105, 145-155.

33. Quapp, W. Computers Math Appl 2001, 41, 407-414.

34. Quapp, W. Optimization 2003, 52, 317-331.

35. Bofill, J. M. J Comput Chem 1994, 15, 1-11.

36. Chu, J.-W.; Trout, B. L.; Brooks, B. R. J Chem Phys 2003, 119, 12708-12717.

37. Miller, W.; Handy N. C.; Adams, J. E. J Chem Phys 1980, 72, 99-112.

38. Sun, J.-Q.; Ruedenberg, K. J Chem Phys 1993, 98, 9707-9714.

39. Quapp, W.; Imig, O.; Heidrich, D. In: Heidrich, D. (ed.) The Reaction Path in Chemistry: Current Approaches and Perspectives; Kluwer: Dordrecht, 1995; Pages 137-160.

Figure Captions

Figure 1. Initial arc of points (see Table I) and IRC (left) on PES (9) using the projector (7). The right part of the chain is destroyed by zigzagging.

Figure 2. RGF to direction $\mathbf{r} = (1, 0)$ on PES (9) using the projector (13). The vector field $\mathbf{p}$ directs every string point of an initial chain to the curve searched for, the solid line.

Figure 3. Three RGFs on PES (9) to different directions $\mathbf{r}$. The long dashed curve is a good reaction path model. Also the full curve may serve as an RP model. TP is a turning point; a trajectory with a TP is not an RP model, see text.

Figure 4. The Müller-Brown potential.[21] The energy is cut at 200 units.

Figure 5. Equipotential lines of the Müller-Brown potential with an initial chain, as well as with the moves of the string method with projector (13), after k=9 loops. Used is the projector to the direction of the initial chain. The true RGF curve is also shown.

Figure 6. The energy over two chains of Ar_7 clusters with LJ_7 potential. The higher one is the straight initial guess between the two minima: a pentagonal bipyramid (left) and a capped octahedron (right), see Table II. The lower chain is the converged result of the string method with projector (13) after $k = 10$ loops.

Figure 7. Gradient extremals on PES (9) and an approximated part of the valley floor GE using the projector (17). The curve represented by black dots is the initial guess. It is the NT of Fig.2. The vector field $\mathbf{p}$ of the projector $-\mathbf{P_w}\ \mathbf{H}\ \mathbf{g}$ excludes a global fit of GEs. Local parts of GEs may be approximated, if the curly field does not turn off. Shown is the chain over the SP after $k = 18$ loops. The accuracy cannot be increased by further loops.

Figure 8. Gradient extremals on PES (9) and an approximated part of the valley floor GE using the projector (24). The curve represented by black dots is the initial chain. The vector field $\mathbf{p}$ of the projector $-\mathbf{P_a}\ \mathbf{g}$ is adapted to the search of the entire valley floor GE. The converged result is reached after $k = 9$ loops.

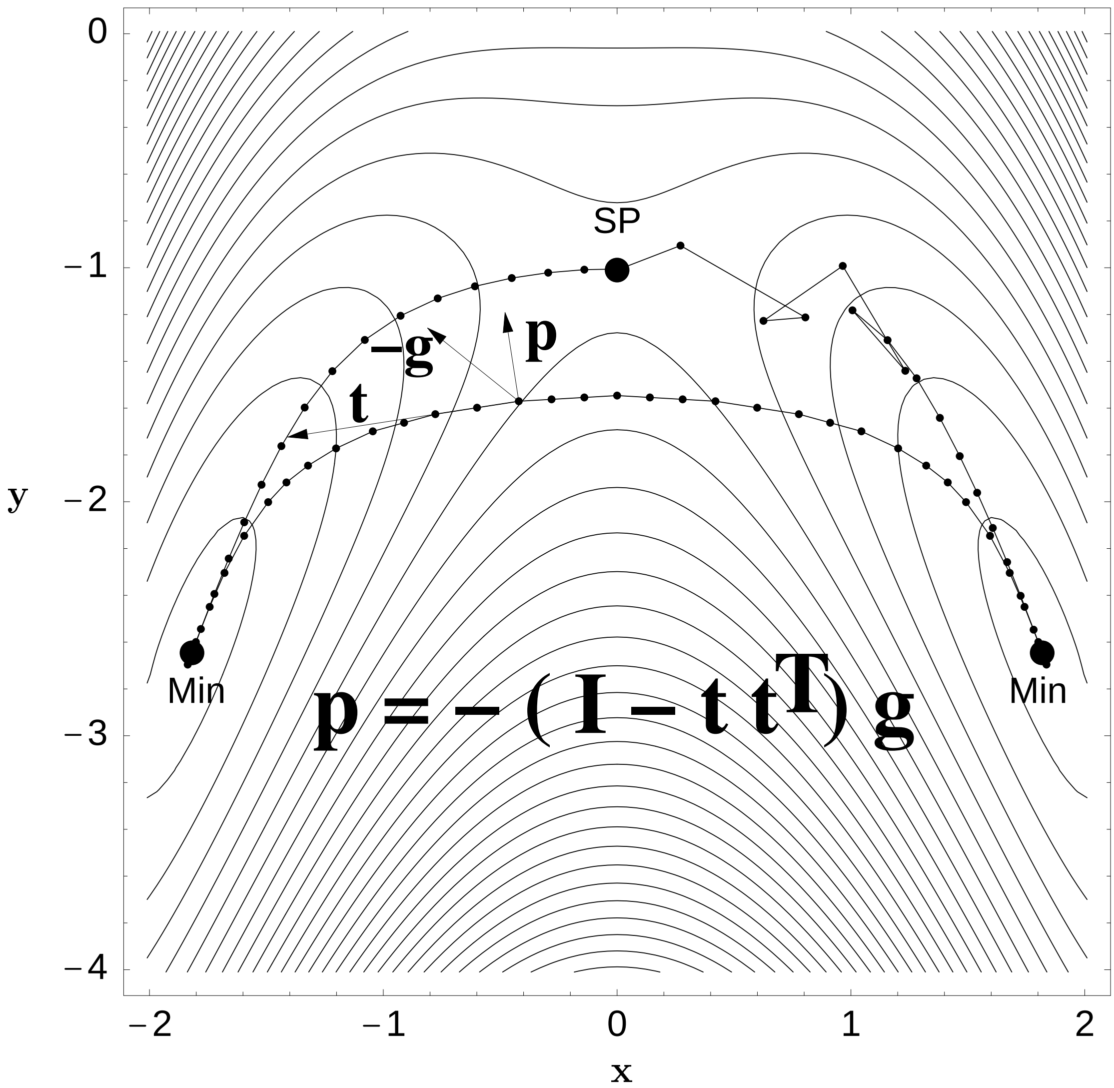

FIGURE 1. W. Quapp

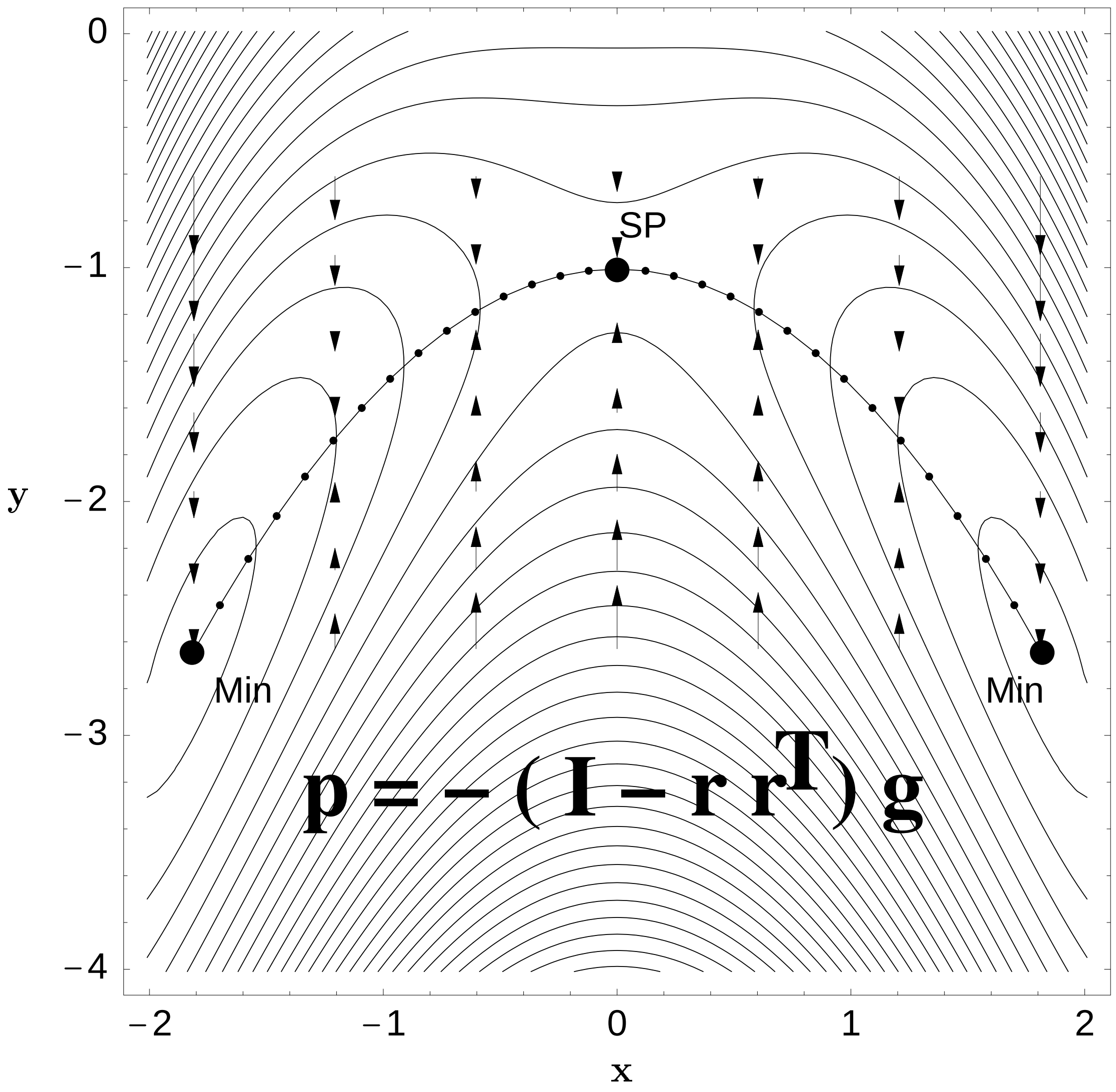

FIGURE 2. W. Quapp

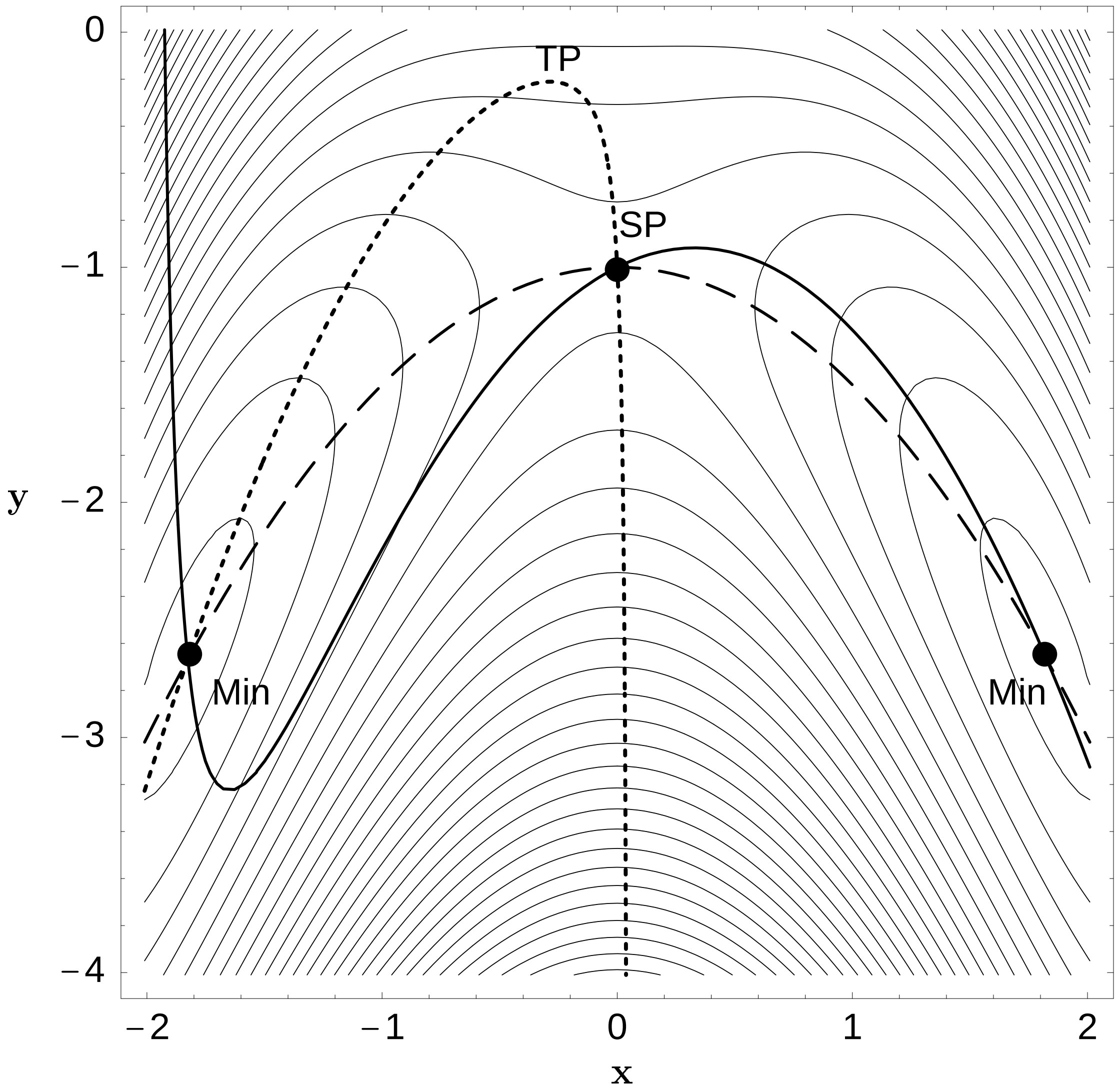

FIGURE 3. W. Quapp

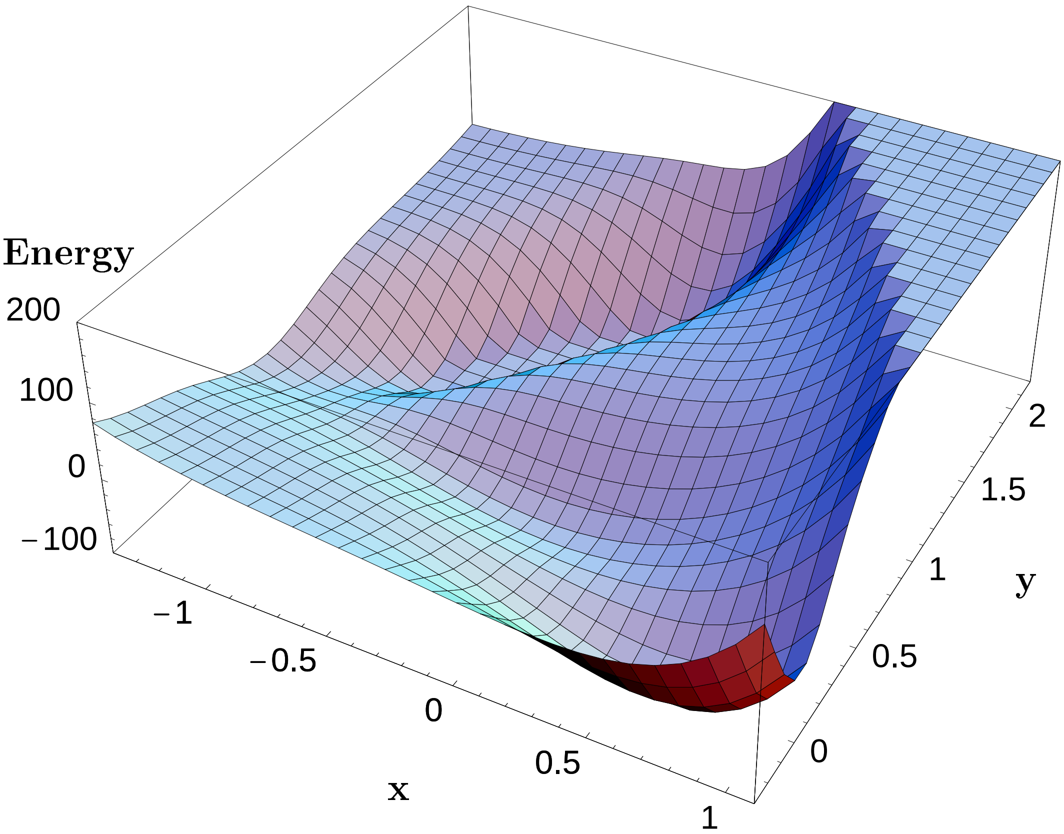

FIGURE 4. W. Quapp

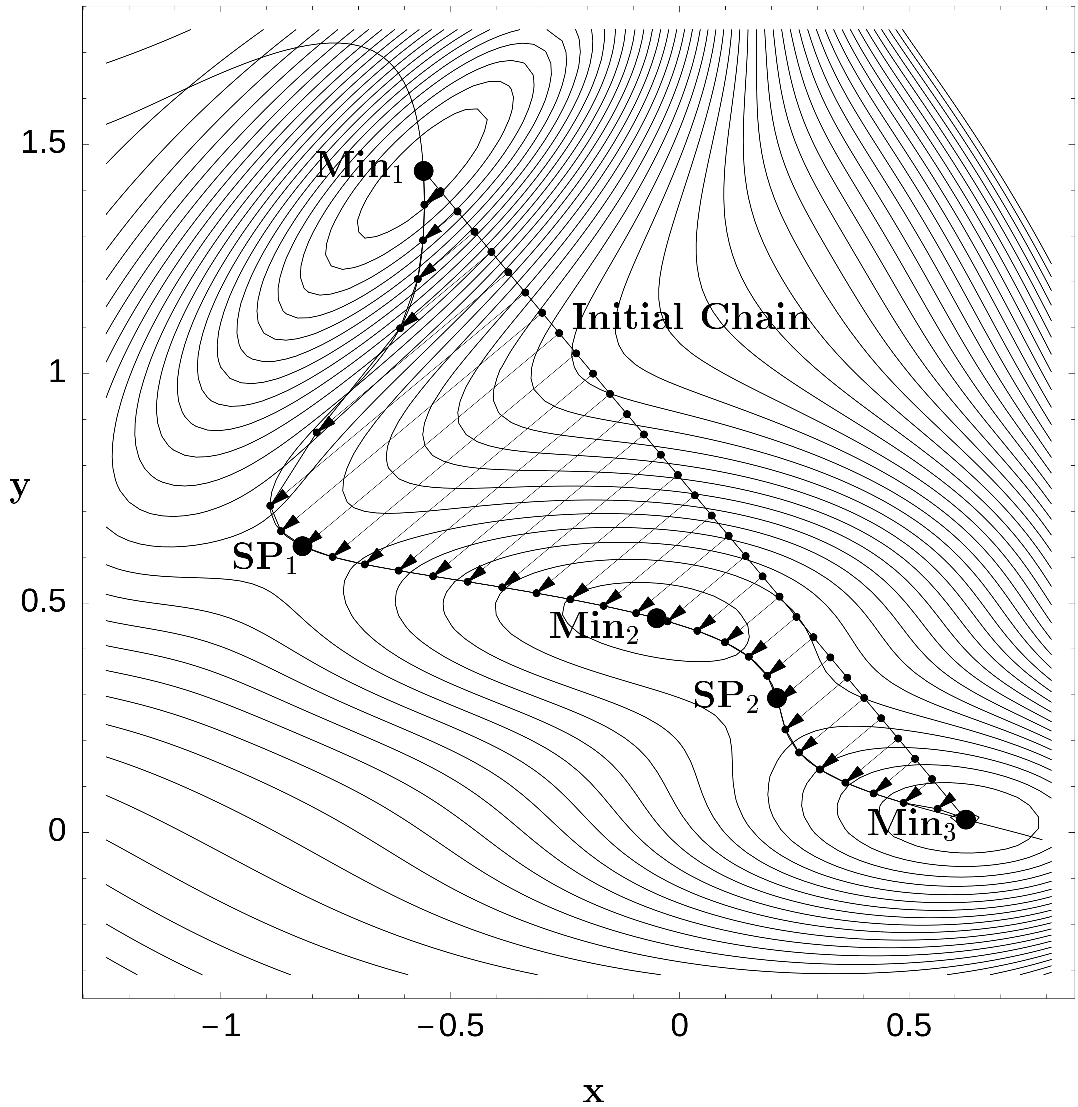

FIGURE 5. W. Quapp

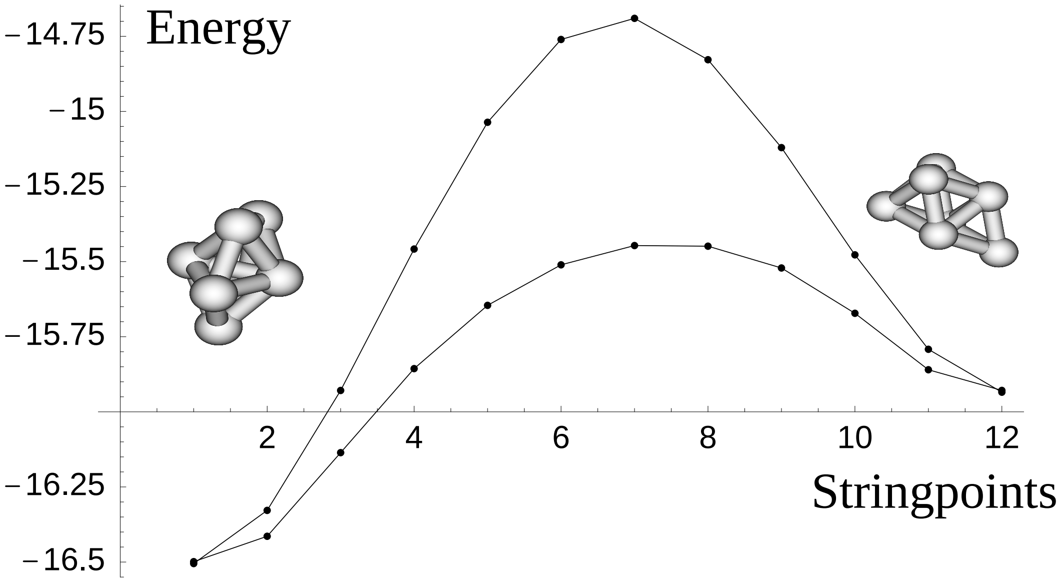

FIGURE 6. W. Quapp

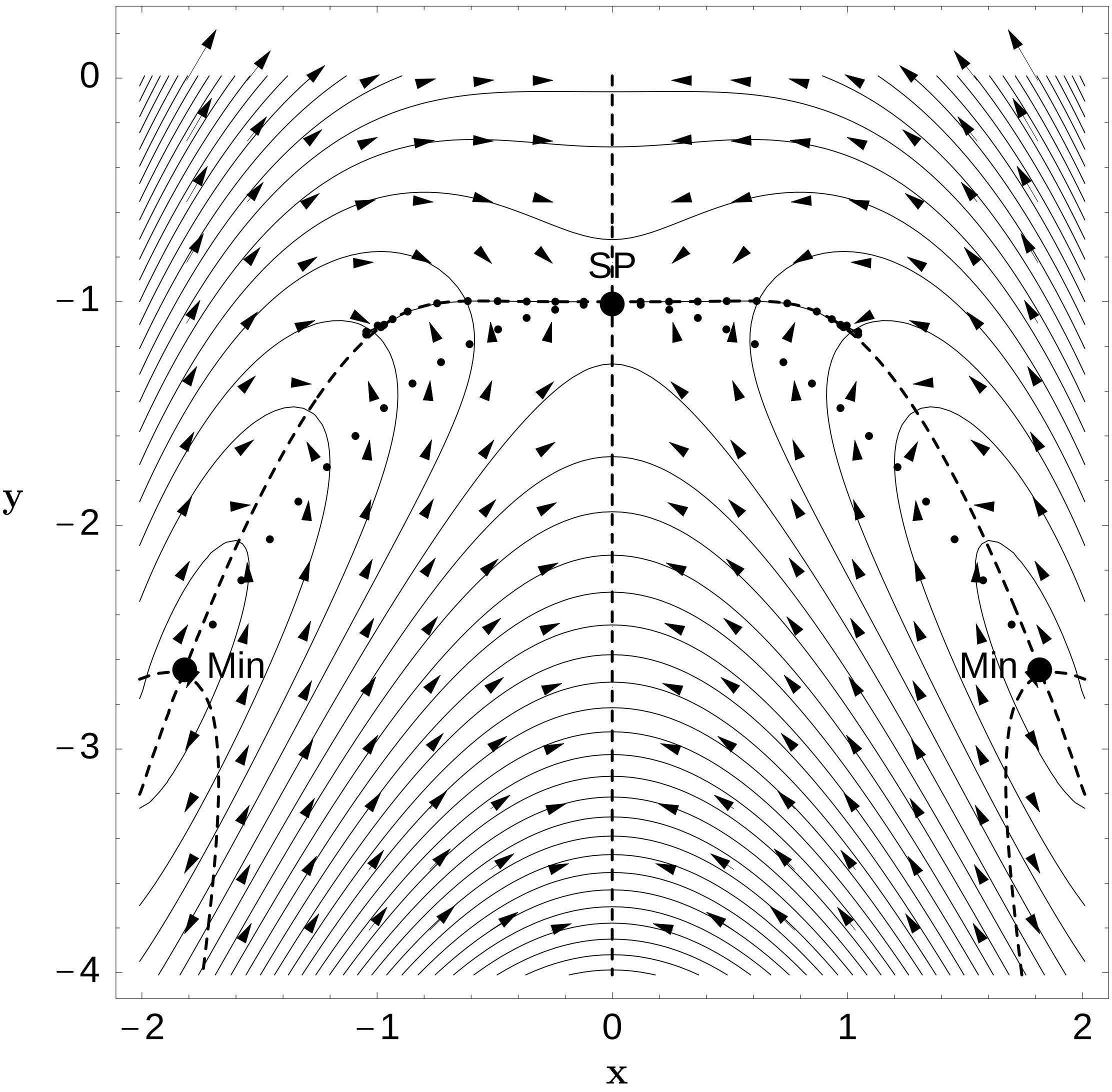

FIGURE 7. W. Quapp

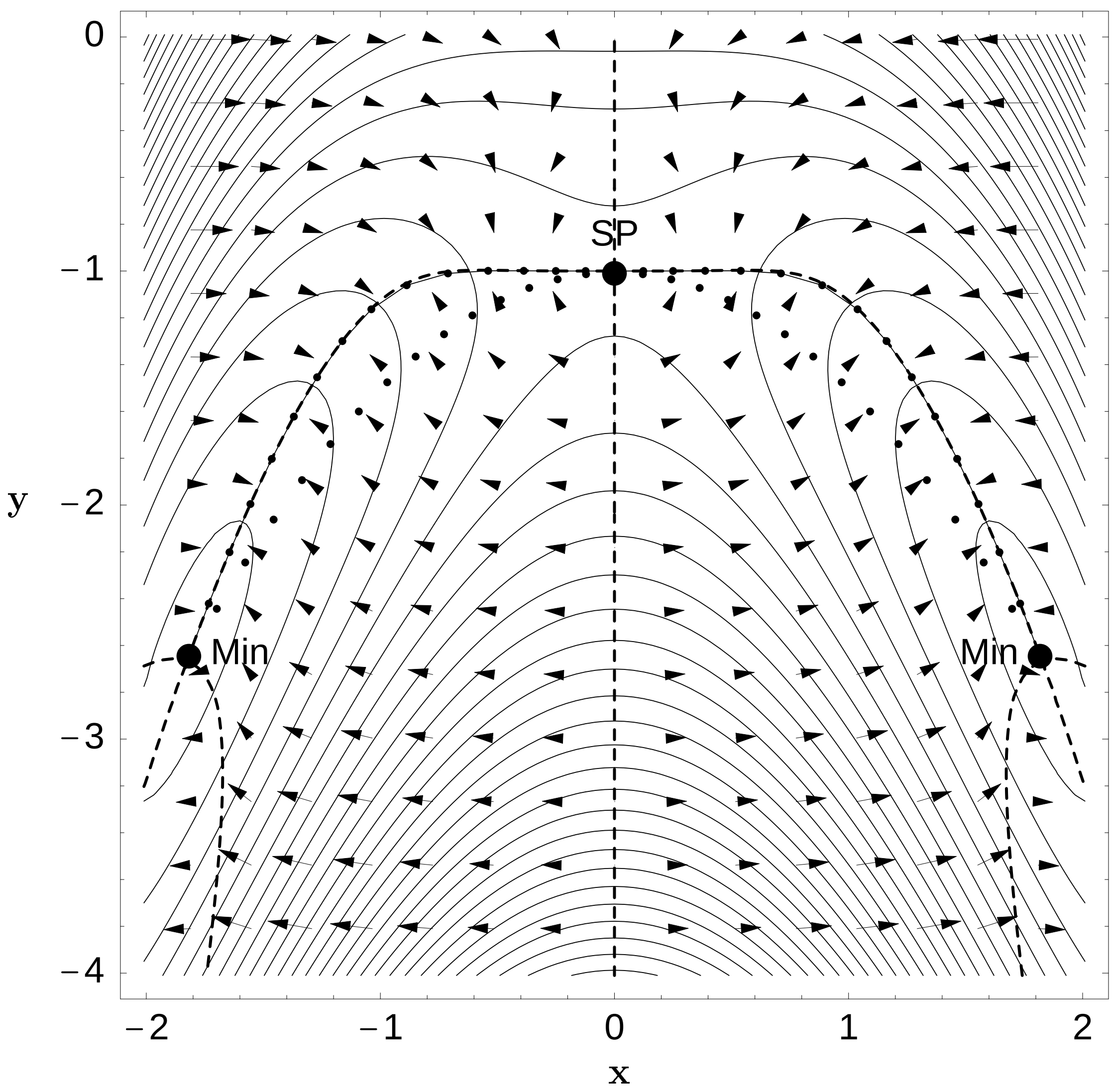

FIGURE 8. W. Quapp

Table I. Initial chain for IRC search

x	y
-1.839286214	-2.710162124
-1.771267766	-2.501291195
-1.714866194	-2.344580505
-1.664800455	-2.223209085
-1.596948599	-2.085363350
-1.522961419	-1.925516704
-1.441353970	-1.757721992
-1.349117833	-1.585488505
-1.244156101	-1.413683637
-1.123669832	-1.251082924
-0.980532930	-1.115237478
-0.804408860	-1.038180507
-0.621268733	-1.015200747
-0.449881793	-1.009989716
-0.286323818	-1.007955973
-0.120011417	-1.007468749
0.000000000	-1.070039364
0.151270908	-1.006854061
0.317831491	-1.007351093
0.484680929	-1.008964616
0.656032141	-1.016935968
0.853329044	-1.049476236
1.027395340	-1.151112497
1.163032705	-1.299870840
1.281313150	-1.471439756
1.380362626	-1.641677392
1.465186852	-1.805297074
1.539633438	-1.961756735
1.606065617	-2.112309187
1.661022796	-2.257258251
1.702256171	-2.373750468
1.761229138	-2.513088454
1.838787446	-2.708460277

Table II. Initial, SP, and end structures of an isomerization path of LJ_7 cluster

Pentagonal Bipyramid

Atom	x	y	z
1	3.547759056	-0.003635069	0.000034068
2	0.298321962	0.002161463	1.948768591
3	1.307957380	3.090755211	0.000034060
4	1.296948975	-3.090042125	0.000034077
5	-2.333937427	-1.903134748	0.000034073
6	-2.327119736	1.916808429	0.000034063
7	0.298321962	0.002161453	-1.948700455

Saddle Point

Atom	x	y	z
1	3.839573630	0.019621762	1.485070022
2	0.119268950	-0.001017963	2.340390978
3	1.543170865	2.558862317	-0.076665641
4	1.568986775	-2.538584348	-0.083724536
5	-2.174049303	-1.916195926	0.009560972
6	-2.193214033	1.896561464	0.015492854
7	0.052078103	0.005695683	-2.393218203

Capped Octahedron

Atom	x	y	z
1	3.786668036	0.037900772	2.552105859
2	-0.016702516	-0.007700255	2.654712299
3	1.791808528	1.925378097	-0.081562280
4	1.837159238	-1.881526182	-0.092408998
5	-1.950108375	-1.921455333	0.017197767
6	-1.995328040	1.874450763	0.028013202
7	-0.146824507	0.006038281	-2.711171021