An Application of the Maximum Principle in Chemistry: A Method to Locate Transition States

Josep Maria Bofill [⋆,1] Wolfgang Quapp [#,2]

October 9, 2016

Keywords: Optimal Control, Transition State, Saddle Point, Reaction Path, Minimum Energy Path, Maximum Principle

Abbreviations: GAD, PES, SP, TP, VRI, MEP, NT, GE, IRC

Subject area: general interest in chemistry

Abstract

The solution curves of the Gentlest Ascent Dynamics (GAD) follow a maximum principle. Under application of the Pontryagin maximum principle, it is demonstrated that the optimal control vector is exactly given by the second GAD equation. The variational nature of GAD curves is discussed, as well as that of some other known reaction pathways.

[1]Universitat de Barcelona, Departament de Química Inorgànica i Orgànica, Secció de Química Orgànica, Universitat de Barcelona, *and* Institut de Química Teòrica i Computacional, Universitat de Barcelona, (IQTCUB), Martí i Franquès, 1, 08028 Barcelona, Spain

⋆email: jmbofill@ub.edu

[2]Mathematisches Institut, Universität Leipzig, PF 100920, D-04009 Leipzig, Germany

#email: quapp@uni-leipzig.de

1 Introduction

An important goal in chemistry is the control of processes, if possible in an optimal manner, for the quantitative conversion of a molecule to a desired product. Normally, the control in chemistry is achieved modifying the thermodynamics of the process or the reaction through the external parameters like temperature, pressure, concentration, or solvent. This is the first and the most widely used control in chemistry. A second way is to manipulate the kinetics of the process by adding or modifying an appropriate catalyst. The third way is the most recent and now under strong development, namely, the control of the reactions through special light sources offering the opportunity to control quantum systems coherently. Within the third option were first the theoretical proposals due to Brumer and Shapiro[1], and to Tannor-Kosloff-Rice[2]. The latter authors proposed a pump-pump scheme where the laser light is used to create and steer nuclear wavepackets to control the molecular reaction. The first experimental realization of the theoretical proposal was demonstrated by Zewail and coworkers[3,4]. From a theoretical point of view, optimal pulses steering a reaction coherently from the given reactant to a desired or predefined product: that can be found in a more direct way by utilizing the approach of optimal control theory (OCT)[5,6].

Other important applications of optimal control theory are given in the area of chemical engineering. The chemical industry moves towards the field of life sciences in which fed-batch processes are predominant (e.g. production of food), optimization and control of fed-batch bioreactors have become more challenging than ever. By programming substrate feeding, one can control important phenomena such as substrate inhibition, glucose effect, and catabolite repression. From the control engineering point of view, fed-batch processes are quite challenging, since the optimization of the substrate feed rate is a dynamic problem[7].

Finally, workers in theoretical chemistry have proposed algorithms to find transition states that are based in the optimal control theory[8]. This theory is contained therein to find the set of rules of a system in a way that a certain optimality criteria is achieved. It is the part of mathematics that formalizes and solves the problem to choose the best way of realizing a controlled process in a prescribed sense. Depending of the parameters, or control parameters, which usually are subject to some constraints, the optimal control process is described

through differential or integral functionals. According to the formulation of the problem, the search of the controls and the realization of the process is chosen in accord to certain prescribed constraints. In fact the term "mathematical theory of optimal control" is applied to the part of mathematical science dealing with the solution of non-classical variational problems of optimal control. The type of problems permits the search of non-smooth functionals and arbitrary constraints related with the control parameters or on the dependent variables. The term covers mathematical methods involving a statistical process or the dynamic optimization, and its interpretation is given in terms of applied procedures for adopting optimal solutions. Taking this into account, the mathematical theory of optimal control contains elements of operations research, mathematical programming, game theory and machine learning. The set of problems studied in the mathematical theory of optimal control have arisen from practical demands like automatic control theory. In methods and in applications, the mathematical theory of optimal control is closely related with analytic mechanics, in the areas relating to the variational principles of classical mechanics.

The variational theory is applied extensively in theoretical chemistry especially in quantum chemistry, see e.g. McWeeny[9], Carbó–Dorca[10] and Bofill[11].

The reaction path (RP) concept[12] is one of the most widely used models in theoretical chemistry. The nature of many types of curves representing RPs has been proved to be variational[8,13–17]. The recently proposed curve for a RP model, the gentlest ascent dynamics[18] is an example of a curve that falls in the group of a variational problem of optimal control[8]. We describe such a reaction path in this chapter. Its nature is based on the Maximum Principle, the basis of the OCT, and it can be used to locate transition states on a Potential Energy Surfaces (PES). Additionally we review the variational nature of some other types of paths. We treat steepest descent, gradient extremal, and distinguished coordinate or its modern version, the Newton trajectory. An extension of the model and its behavior is also discussed.

2 The Maximum Principle as a Basis of the Gentlest Ascent Dynamics Model

In this section we report a proof of the Optimal Control character of the curves of the Gentlest Ascent Dynamics (GAD). For this purpose we consider a system with N degrees of freedom represented by a point vector $\mathbf{x} \in \mathbb{R}^N$. Curves in the $\mathbb{R}^N$ are usually characterized by $\mathbf{x}(t)$ with a parameter t. The potential energy is described by the PES function, $V(\mathbf{x})$. The concept of the GAD model is that of a dynamical system. The solution curves of the GAD equations[18] evolve from a point close to a minimum to an stationary point (SP) on the PES. The GAD model is based on the gradient field of the PES, $\mathbf{g}(\mathbf{x}) = \nabla_{\mathbf{x}} V(\mathbf{x})$, and a normalized control vector, $\mathbf{w}$. The Hessian matrix is also used, $\mathbf{H}(\mathbf{x}) = \nabla_{\mathbf{x}} \mathbf{g}^T(\mathbf{x})$. The control vector itself is generated on the path, point by point, by a continuous version of the power method for finding the eigenvector of the Hessian matrix which belongs to the smallest eigenvalue. The first GAD equation for the tangent or velocity vector $\dot{\mathbf{x}}$ of a GAD curve is the sum of the reverse (negative) gradient plus two times an effect of the control vector, $\mathbf{w}$, shorten with the projection on the gradient

$$\dot{\mathbf{x}} = -\left[\mathbf{I} - 2\mathbf{w}\mathbf{w}^T\right] \mathbf{g}(\mathbf{x}) \ , \tag{1}$$

where we assume that the $\mathbf{w}$-vector is normalized. Geometrically, the matrix $\left[\mathbf{I} - 2\mathbf{w}\mathbf{w}^T\right]$ is a mirror transformation at the mirror line through the control vector $\mathbf{w}$, the Householder orthogonal transformation[19]. Note that $(\mathbf{w}\mathbf{w}^T)$ is a dyadic product matrix. The control vector $\mathbf{w}(t)$ depends on the curve parameter, t. Finding the variational bases of this model is, in general, difficult. It was proved to be an optimal control problem by Bofill and Quapp[8] based on Zermelo's navigation problem[20], see also Zermelo's navigation problem in Carathéodory's book[21] and Carathéodory's 1926 article[22] which can be seen as precursors of the maximum principle and OCT. They are the foundations attributed to the field of variational studies which are realized during the last fifty years of the 20th century. In the GAD model is (realistically) assumed that the gradient vector field cannot be controlled and that the control is to execute by the normalized vector, $\mathbf{w}$, which is here generated by the power method to find the eigenvector with the lowest eigenvalue of the Hessian, $\mathbf{H}$; thus

$$\dot{\mathbf{w}} = -\left[\mathbf{I} - \mathbf{w}\mathbf{w}^T\right] \mathbf{H}\mathbf{w} \ , \tag{2}$$

the matrix $\left[\mathbf{I} - \mathbf{w}\mathbf{w}^T\right]$ is the projection orthogonal to the control vector, $\mathbf{w}$.

The equivalence between Zermelo's navigation problem and the GAD system can be seen as follows: in the navigation problem the central question is the present location of a ship in the sea, with a given current distribution characterized by a local dependent vector field. The current is assumed to be independent of a time, it only depends on the position. One desires to find the optimal control of the ship so to reach the destination in the shortest possible time. In the GAD model the gradient vector field of the PES function can be thought of as representing the current of the sea, which we cannot change, whereas the normalized vector $\mathbf{w}$ determines the control. The destination is the next SP of the PES. We recall that the set of coupled first-order ordinary differential equations, Eq.(1) and (2), constitute the fundamental expressions of the GAD model[18,23] In the recent reference[8] we use both, a device due to Zermelo as well as the Lagrange multipliers method,[20] see also Carathéodory[21]. There the variational nature of the GAD model was proved. Now we will proof that GAD is an example of OCT based on a Legendre construction.

2.1 Variational Necessary Conditions

Let us consider a controlled object which is represented by a point $\mathbf{x} = (x_1, \cdots, x_N)^T$ in the N-dimensional configuration space, and we use the system of N non-autonomous differential equations (1) with the control parameters $\mathbf{w} = (w_1, \cdots, w_N)^T$ which are the components of the $\mathbf{w}$-vector. The values of the $\mathbf{w}$-vector are assumed to be on the unit sphere of $\mathbb{R}^N$ $\mathbf{w}^T\mathbf{w} = 1$. For the reason we have $N-1$ control parameters in the GAD model. Now, we have given an initial state that is supposed to be a minimum of the PES, $\mathbf{x}_0 = \mathbf{x}(t_0)$, and a final state, a stationary point of index one of this PES, $\mathbf{x}_{TS} = \mathbf{x}(t_f)$ with a variable t_f depending from the pathway to the TS. We will find a control $\mathbf{w}(t)$-vector on the unit sphere of $\mathbb{R}^N$ for $t_0 \leq t \leq t_f$ such that it minimizes the transition of the state point $\mathbf{x}$ moving on a GAD path from $\mathbf{x}_0$ to $\mathbf{x}_{TS}$ according to the non-autonomous system of Eq.(1).

In a more precise way, the GAD model consists in the determination of the minimum of the t-parameter, $J\left[\mathbf{x}_{TS}(\mathbf{w}(t_f))\right] = t_f - t_0$. A controlled point can be evolved from a given minimum point of the PES, $\mathbf{x}_0 = \mathbf{x}(t_0)$, to a final transition state of this PES, $\mathbf{x}_{TS} = \mathbf{x}(t_f)$. The evolution of the test point is described by the system of ordinary differential equations (1). Note that $\mathbf{x}(t)$ is a N-dimensional vector of the configuration space, where $\mathbf{w}(t)$ is a normal-

ized N-dimension vector of the control parameters. Due to the normalization condition of the $\mathbf{w}(t)$-vector we have $N-1$ control parameters, and for every t these control parameters belong to the unit sphere.

Since t_0 is fixed, the required minimum t_f is merely the minimization of the functional $J\left[\mathbf{x}_{TS}(\mathbf{w}(t_f))\right]$ that depends on the chosen $\mathbf{w}(t)$-control normalized vector. Thus the GAD model is a t-parameter-optimal control problem and can be considered as a particular instance of the Mayer problem of the Theory of calculus of variations, and is obtained from these problems by the special form of the functional to be optimized. The GAD, as a case of an optimal control problem, must satisfy the Pontryagin Maximum Principle, which is a necessary condition that generalizes the necessary conditions of Euler and Weierstrass, used in the classical Theory of calculus of variations[21,24], see here the Mayer problem. From this formulation of the GAD model and following Pontryagin[25] we can formulate the GAD problem as follows: if the pair of vectors $\mathbf{x}(t)$ and $\mathbf{w}(t)$ for $t_0 \leq t \leq t_f$ is an optimal solution then there exists a nonzero covector-function $\mathbf{y}(t)$ such that $\mathbf{x}(t)$, $\mathbf{w}(t)$ and $\mathbf{y}(t)$ for $t_0 \leq t \leq t_f$ is a solution to the system of differential equations Eq.(1) and the equation

$$\dot{\mathbf{y}}(t) = \frac{d}{dt}\mathbf{y}(t) = -\left[\nabla_{\mathbf{x}}\dot{\mathbf{x}}(t)^T\right]\mathbf{y}(t) = \mathbf{H}(\mathbf{x}(t))\left[\mathbf{I} - 2\mathbf{w}(t)\mathbf{w}(t)^T\right]\mathbf{y}(t)\ , \quad (3)$$

where $\dot{\mathbf{x}}(t)$ is that given by Eq.(1), and along the solution, for every t a type of maximization with respect to the normalized $\mathbf{w}(t)$-vector is satisfied. The type of maximization will be treated below.

We suppose that the set of admissible $(w_1, \cdots, w_N)$ values of the control belongs to the unit sphere with $\mathbf{w}^T\mathbf{w} = 1$. Since the curve, $\mathbf{x}(t)$, for $t_0 \leq t \leq t_f$ is optimal, it traverses at each value of t a plane. The covector-function, $\mathbf{y}(t)$, is orthogonal to this plane, and Eq.(3) represents the "transportation" of this plane along the optimal curve $\mathbf{x}(t)$, for $t_0 \leq t \leq t_f$. The first order variation on $\mathbf{w}(t_0)$ produces a first order variation on the difference vector, $\mathbf{x}_v(t_f) - \mathbf{x}(t_f)$, being $\mathbf{x}_v(t)$ the curve that starts at $\mathbf{x}_0 = \mathbf{x}_v(t_0) = \mathbf{x}(t_0)$ but with a different $\mathbf{w}(t_0)$-vector of the control. According to Pontryagin et al. the above difference vector $\mathbf{x}_v(t_f) - \mathbf{x}(t_f)$ is orthogonal to the $\mathbf{y}(t_f)$-vector and all the difference vectors generated in this way are orthogonal to the $\mathbf{y}(t_f)$-vector. The full set of difference vectors are in a plane whose normal is the $\mathbf{y}(t_f)$-vector and which contains the point $\mathbf{x}(t_f)$. The plane formula is $\mathbf{y}(t_f)^T\left[\mathbf{x}_v(t_f) - \mathbf{x}(t_f)\right] = 0$ and the dimension of the subspace formed by the set of difference vectors,

$\{[\mathbf{x}_v(t_f) - \mathbf{x}(t_f)]\}$, is $N-1$.

Now a problem is to solve. We have only the initial point $\mathbf{x}_0 = \mathbf{x}(t_0)$, and we take an arbitrary initial value $\mathbf{y}_0 = \mathbf{y}(t_0) \neq \mathbf{0}$, and we attempt to solve the system of $2N$ equations, namely, Eq.(1) and (3) with $3N$ unknowns, $\mathbf{x}(t)$, $\mathbf{y}(t)$, and $\mathbf{w}(t)$ proceeding along an arbitrary extremal GAD curve passing through $\mathbf{x}_0$. Clearly we need another equation for $\mathbf{w}(t)$ such that the above problem can be solved uniquely. If this is possible then the $2N$ unknown parameters are left, $\mathbf{x}(t)$ and $\mathbf{y}(t)$, subject to the system of $2N$ differential Eq.(1) and (3) and the initial conditions, $\mathbf{x}_0 = \mathbf{x}(t_0)$ and $\mathbf{y}_0 = \mathbf{y}(t_0)$. Because the adjoint Eq.(3) is linear in $\mathbf{y}$ the function $\mathbf{y}(t)$ is defined up to a nonzero constant factor. This property we will use later.

Collecting the two necessary conditions (1) and (3) given above we can put a certain combination of symbols in the scalar-valued function of three arguments

$$\mathbb{H}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{w}(t)) = \dot{\mathbf{x}}(t)^T \mathbf{y}(t) = -\mathbf{g}(\mathbf{x}(t))^T \left[\mathbf{I} - 2\mathbf{w}(t)\mathbf{w}(t)^T\right] \mathbf{y}(t) \qquad (4)$$

where Eq.(1) has been used. The function enables us to rewrite the system of equations (1) and (3) as a Hamiltonian system and the function of Eq.(4) as a Hamiltonian function,

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \nabla_{\mathbf{y}} \mathbb{H}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{w}(t)) \ , \\ \dot{\mathbf{y}}(t) &= -\nabla_{\mathbf{x}} \mathbb{H}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{w}(t)) \ . \end{aligned} \qquad (5)$$

We have used that the Hamiltonian function is a scalar function, being a product of a matrix that is multiplied by two different vectors from the left and from the right. For this reason this scalar function can be written as

$$\begin{aligned} \mathbb{H}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{w}(t)) &= -\mathbf{y}(t)^T \left[\mathbf{I} - 2\mathbf{w}(t)\mathbf{w}(t)^T\right] \mathbf{g}(\mathbf{x}(t)) \\ &= -\mathbf{g}(\mathbf{x}(t))^T \left[\mathbf{I} - 2\mathbf{w}(t)\mathbf{w}(t)^T\right] \mathbf{y}(t) \ . \end{aligned}$$

So, $\mathbb{H}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{w}(t))$ is a function of $\mathbf{x}$ through the gradient vector, $\mathbf{g}(\mathbf{x}(t))$. Applying Eq.(5b) to this Hamiltonian function we get Eq.(3).

The Hamiltonian function of Eq.(4) is the GAD Hamiltonian, see also Eqs.(18) and (41) of Ref. 8. Because the optimal GAD curve, $\mathbf{x}(t)$, traverses at this point a plane, (it is not tangent to this plane) to which the $\mathbf{y}(t)$-covector is the normal vector, we obtain

$$\mathbb{H}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{w}(t)) = \mathbf{y}(t)^T \dot{\mathbf{x}}(t) > 0 \ . \qquad (6)$$

Hence the plane divides the configuration space, $\mathbb{R}^N$, in two distinguishable half-spaces, $\mathbb{R}^N_-$, before the optimal GAD curve, $\mathbf{x}(t)$, intersects the plane, and $\mathbb{R}^N_+$, after the intersection. In particular every initial control vector, $\mathbf{w}_v(t_0)$, different with respect to the initial optimal control vector $\mathbf{w}(t_0)$, displaces the endpoint of the optimal curve $\mathbf{x}(t_f)$ to another point. The real displacement $\Delta\mathbf{x}(t_f) = \mathbf{x}_v(t_f) - \mathbf{x}(t_f)$ is certainly a nonlinear function. Generally, it stays off the plane, namely, $\mathbf{x}(t_f) + \Delta\mathbf{x}(t_f) \in \mathbb{R}^N_-$ or $\mathbf{x}(t_f) + \Delta\mathbf{x}(t_f) \in \mathbb{R}^N_+$. From a geometric point of view, the optimal GAD curve, $\mathbf{x}(t)$, $t_0 \leq t \leq t_f$ consists in the assertion that the displacement of its endpoint due to the differences of the control vector, $\mathbf{w}_v(t)$ and $\mathbf{w}(t)$ falls into the half-space $\mathbb{R}^N_-$, in other words, $\mathbf{x}(t_f) + \Delta\mathbf{x}(t_f) \in \mathbb{R}^N_-$ for any variation of the control vector.

Along the above reasoning we can conclude that additional to the necessary conditions on a GAD curve given by the Eqs.(1) and (3) another necessary condition emerges. The scalar product $\mathbf{y}(t_f)^T\Delta\mathbf{x}(t_f)$, where $\Delta\mathbf{x}(t_f)$ is obtained from any variation of the initial control vector is nonpositive provided that the covector $\mathbf{y}(t_f)$, which is the normal to the plane, is correctly normalized (directed toward the half-space $\mathbb{R}^N_+$). In other words, $\mathbf{y}(t_f)^T\Delta\mathbf{x}(t_f) \leq 0$, for any variation of the initial control vector, see Fig.(1).

In Fig.(1) we show the optimal GAD curve, $\mathbf{x}(t)$, which starts at $\mathbf{x}_0 = \mathbf{x}(t_0)$ with the initial control vector, $\mathbf{w}(t_0)$. At $t = t_f$ the curve transverses the hyperplane of dimension $N-1$. The normal of this hyperplane is the $\mathbf{y}(t_f)$-vector. The other curve $\mathbf{x}_v(t)$ with initial control vector $\mathbf{w}_v(t_0)$ does not achieve the $\mathbf{x}_{TS}$ position at the time $t = t_f$. It is not the optimal curve and for this reason $\mathbf{y}(t_f)^T\Delta\mathbf{x}(t_f) = \mathbf{y}(t_f)^T\left[\mathbf{x}_v(t_f) - \mathbf{x}(t_f)\right] \leq 0$, which is the additional necessary condition supporting the Maximum Principle. In the GAD model this additional necessary condition is achieved by maximization of the Hamiltonian given in Eq.(4) with respect to any direction $\mathbf{y}(t)$ under the normalized control vector $\mathbf{w}(t)$.

For the GAD model this necessary condition can be written as

$$\begin{aligned}\mathbb{H}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{w}(t)) &= \max_{\{\mathbf{y}(t)|\mathbf{y}^T\mathbf{y}=1\}} \mathbb{H}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{w}(t)) \\ &= \max_{\{\mathbf{y}(t)|\mathbf{y}^T\mathbf{y}=1\}} \{-\mathbf{y}(t)^T[\mathbf{I} - 2\mathbf{w}(t)\mathbf{w}(t)^T]\mathbf{g}(\mathbf{x}(t))\} \,, \end{aligned} \tag{7}$$

where Eq.(4) has been used. The direction $\mathbf{y}(t)$ that maximizes the Hamilto-

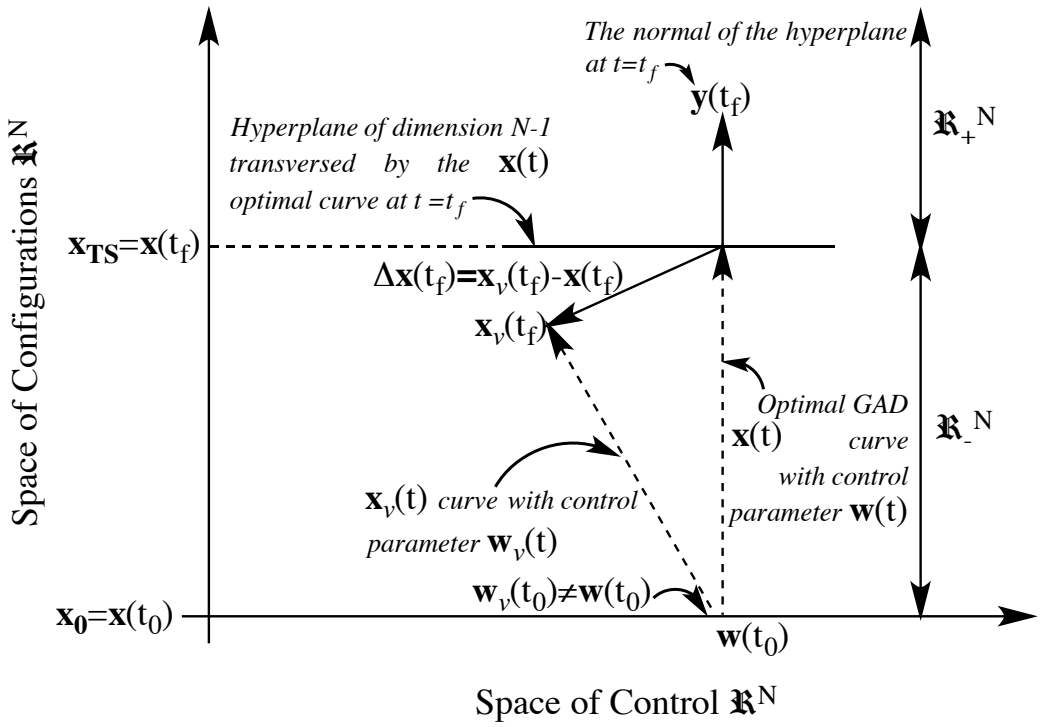

Figure 1: Scheme of the Pontryagin Maximum Principle applied to the GAD model.

nian is attained as follows:

We observe that $\mathbf{w}^T(t)$ is a left eigenvector of the matrix $[-\mathbf{I}+2\mathbf{w}(t)\mathbf{w}(t)^T]$ with eigenvalue $+1$. This is the only one positive eigenvalue. All the possible $(N-1)$ linear independent eigenvectors orthogonal to $\mathbf{w}(t)$ have the $(N-1)$-fold degenerate eigenvalue -1. Thus to get a maximum in Eq.7 we must chose the direction $\mathbf{y}(t) = \mathbf{w}(t)$. Taking this direction we would get in Eq.(4) the value $\mathbf{w}(t)^T\mathbf{g}(\mathbf{x}(t))$ which still depends on the length of the gradient, the $\mathbf{g}(\mathbf{x}(t))$-vector. To avoid this and to ensure that for any $t_0 \leq t \leq t_f$ the Hamiltonian function of Eq.(4) achieves a constant maximum value, we take

$$\mathbf{y}(t) = \frac{\mathbf{w}(t)}{\mathbf{w}(t)^T\mathbf{g}(\mathbf{x}(t))} \tag{8}$$

for a solution of the maximum direction in Eq.7. With this solution we get

$\mathbb{H}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{w}(t)) = 1$. This is the maximum value which the GAD Hamiltonian achieves along the optimal GAD curve.
Substituting solution (8) in Eq.(3) we obtain a new expression for the $\dot{\mathbf{y}}(t)$-vector,

$$\dot{\mathbf{y}}(t) = -\frac{\mathbf{H}(\mathbf{x}(t))\mathbf{w}(t)}{\mathbf{w}(t)^T\mathbf{g}(\mathbf{x}(t))} \ . \tag{9}$$

If we differentiate Eq.(8) with respect to t and if we equating the resulting expression to Eq.(9) we obtain

$$\frac{\mathbf{w}(t)^T\mathbf{g}(\mathbf{x}(t))\dot{\mathbf{w}}(t) - \mathbf{w}(t)\left[\dot{\mathbf{w}}(t)^T\mathbf{g}(\mathbf{x}(t)) + \mathbf{w}(t)^T\dot{\mathbf{g}}(\mathbf{x}(t))\right]}{(\mathbf{w}(t)^T\mathbf{g}(\mathbf{x}(t)))^2} = -\frac{\mathbf{H}(\mathbf{x}(t))\mathbf{w}(t)}{\mathbf{w}(t)^T\mathbf{g}(\mathbf{x}(t))} \ . \tag{10}$$

Multiplying Eq.(10) from the left by $\left[\mathbf{I} - \mathbf{w}(t)\mathbf{w}(t)^T\right]$ and using that $\mathbf{w}(t)^T\dot{\mathbf{w}}(t) = 0$ and $\left[\mathbf{I} - \mathbf{w}(t)\mathbf{w}(t)^T\right]\mathbf{w}(t) = 0$ we obtain the Eq.(2) which is the searched second GAD equation for $t_0 \leq t \leq t_f$.

With these results we can paraphrase the Pontryagin Maximum Principle[25] applied to the GAD curve model. We assert that the extremal GAD curves are solution of the Hamiltonian system of Eq.(5) or equivalently, Eq.(1)–(2) and according to the Eq.(7) their points maximize the Hamiltonian function (4) with respect to the control vector $\mathbf{w}(t)$; furthermore according to the Eq.(7), along the GAD extremal curves the control vector for which the Hamiltonian function attains its local maximum is given by Eq.(8). Along the optimal GAD curve the Hamiltonian $\mathbb{H}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{w}(t))$ achieves its maximum value being equal to 1.

2.2 Extensions of GAD model

The results of the previous section motivate us to propose new curves based on the GAD model and the requirements of the Pontryagin Maximum Principle. With this consideration a curve was proposed such that the tangent is given by the general expression[8]

$$\dot{\mathbf{x}} = -\mathbf{g}(\mathbf{x}) + f(\phi, \mathbf{x}, \mathbf{w})\mathbf{w} \ , \tag{11}$$

where the control vector, $\mathbf{w}$, is assumed to be normalized. The function $f(\phi, \mathbf{x}, \mathbf{w})$ is a continuous and differentiable function with respect to $\mathbf{x}$, and

ϕ is a constant. In this case the expression for $\dot{\mathbf{w}}$ is

$$\dot{\mathbf{w}} = -\left[\mathbf{I} - \mathbf{w}\mathbf{w}^T\right]\left[\nabla_{\mathbf{x}} f\left(\phi, \mathbf{x}, \mathbf{w}\right) - \mathbf{H}\left(\mathbf{x}\right)\mathbf{w}\right] \ . \tag{12}$$

To delimit the function $f(\phi, \mathbf{x}, \mathbf{w})$, we should organize that if the control vector, $\mathbf{w}$, is associated with the uphill direction of the evolution of the curve, then the general expression should be one that minimizes the potential energy in the subspace orthogonal to the $\mathbf{w}$-vector and maximizes it along the $\mathbf{w}$ direction, like in the GAD case. An interesting special case is

$$\dot{\mathbf{x}} = -\mathbf{g}(\mathbf{x}) + \phi(\mathbf{w}^T\mathbf{g}(\mathbf{x}))\mathbf{w} \ . \tag{13}$$

Here it is $f(\phi, \mathbf{x}, \mathbf{w}) = \phi(\mathbf{w}^T\mathbf{g}(\mathbf{x}))$ with ϕ is greater than one. Taking $\phi = 2$ we obtain the standard GAD model. The behavior of this kind of curves given in Eq.(13) is reported in Ref. 8. Notice that if one takes $f(\phi, \mathbf{x}, \mathbf{w}) = \phi$ being ϕ a positive constant and changing $-\mathbf{g}(\mathbf{x})$ by $\mathbf{g}(\mathbf{x})$ we obtain the original Zermelo problem, namely, $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}) + \phi\mathbf{w}$, where as before it is $\mathbf{w}^T\mathbf{w} = 1$. It is easy to prove that in this case the expression for $\dot{\mathbf{w}}$ coincides with that given in Eq.(2). This possibility has not been yet investigated as a possible curve model to locate transition states, as well as for the representation of reaction paths.

3 The Variational Nature of others Reaction Paths

The reaction path concept is based on the definition of a curve located on the PES, which is monotonely increasing in the potential energy from the reactant minimum to the SP and monotonely decreasing from this point downhill to the product minimum. Many geodesic curves satisfy the definition, and for this reason, there is a large set of curves proposed as a model of the reaction path. The most widely used curves for this purpose are the steepest descent[26–28], the gradient extremals[29–32], and the distinguished reaction coordinate[33], and in its newer version, the Newton trajectory (NT)[34,35].

The steepest-descent/ascent curve is the curve that at each point follows the gradient of the PES. This curve is variational and extremalizes the integral functional[13,14]

$$I(\mathbf{x}) = \int_{t_0}^{t} F(\mathbf{x}, \dot{\mathbf{x}}) = \int_{t_0}^{t} \sqrt{\mathbf{g}^T\mathbf{g}}\sqrt{\dot{\mathbf{x}}^T\dot{\mathbf{x}}}dt' = \int_{s_0}^{s} \sqrt{\mathbf{g}^T\mathbf{g}}ds' \ , \tag{14}$$

where s is the arc length and, $F(\mathbf{x}, \dot{\mathbf{x}})$, is a functional homogeneous of degree one with respect to the argument $\dot{\mathbf{x}}$. The second variation indicates that the

steepest-descent/ascent curve that joints two minimums of the PES through an SP only minimizes the integral functional of Eq.(14). The special curve, the so-called intrinsic reaction coordinate (IRC)[26,27], always satisfies the reaction path definition, but the second variation[13] does not imply that the curve is fully located in a deep valley; see a counter example elsewhere[36]. The second variation is not related to the Minimum Energy Path (MEP) condition. The gradient curve of the type IRC is always a reaction path but it can be or not be a MEP. This depends on the shape of the PES[8].

Another type of curve is the so-called gradient extremal (GE)[37]. A GE curve was proposed as a reaction path some time ago[29,30]. Its definition can be described assuming first that we are on a "valley ground" of the PES with respect to the variations of $\mathbf{x}$ within the equipotential hypersurface, $V(\mathbf{x}) = \nu = const.$ The functional integral to be extremalized is

$$I(\mathbf{x}) = \frac{1}{2}\int_{t_0}^{t} \mathbf{g}\ (\mathbf{x}(t'))^T \mathbf{g}\ (\mathbf{x}(t')) - \lambda(t')[V(\mathbf{x}(t')) - \nu(t')]dt'\ . \qquad (15)$$

The curve that extremalizes the integral functional of Eq.(15) corresponds to the curve that also satisfies at each point the eigenvalue equation

$$\mathbf{H}(\mathbf{x}(t))\ \mathbf{g}(\mathbf{x}(t)) = \lambda(t)\mathbf{g}(\mathbf{x}(t))\ . \qquad (16)$$

This curve is the GE. In Eq.(16) the λ is the eigenvalue and the gradient $\mathbf{g}$ is the corresponding eigenvector. The demonstration of the variational nature of this type of curves was formulated in Ref. 17 (see also references therein). Within the theory of the calculus of variations, the GE problem is classified as a Bolza variational problem which is related to a Lagrangian multipliers problem[21,38]. As mentioned above the GE curve extremalizes the integral functional given in Eq.(15), and in its evolution it transverses the set of equipotential surfaces, $V(\mathbf{x}) - \nu = 0$. From a variational point of view it is important to consider that if the GE curve joins two minimums of the PES and if it does not have on this subarc turning points, then this GE curve describes a reaction path belonging to the category of the MEP[17]. A GE curve joining two minimums of the PES minimizes the integral functional of Eq.(15) if and only if it does not have on this subarc turning points. Otherwise other arbitrary curves joining the same two minimums lower the value of the integral functional of Eq.(15)[17]. We conclude that GE curves which join two minimums of the PES and which do not have turning points in between, minimize the functional given in Eq.(15), and

satisfy the reaction path definition and the MEP requirement. In this case, the model of GE curves has a direct relation between variationality, reaction path and MEP definition. Unfortunately, the GE curves do not cover all the PES; in other words, a GE curve does not exist at the most points of the PES. Additionally, there can be parts of a reaction valley of a PES where no continuous GE exists[39,40].

The distinguished reaction coordinate[33], or its new reformulation, the NT[34,35], is a model curve which is often used to locate SPs. The curve can be used as representation of a reaction path, again if no turning point emerges. The variational nature of the curves was studied in Ref. 16. It corresponds to a problem where the functional only depends on the arguments, coordinates, and the parameter that characterizes the curve

$$I(\mathbf{x}_v) = \int_{x_{rc}^0}^{x_{rc}} V(\mathbf{x}_v, x'_{rc}) dx'_{rc} \ , \tag{17}$$

where the $\mathbf{x}_v$-vector is the coordinate vector $\mathbf{x}$ without the x_{rc} component. It can be shown[16,41] that the curve which extremalizes the functional integral of Eq.(17) is the curve which satisfies the Branin equation[42]

$$\dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt} = \pm \mathbf{A}(\mathbf{x})\mathbf{g}(\mathbf{x}) \ , \tag{18}$$

where $\mathbf{A}(\mathbf{x})$ is the adjoint matrix of the Hessian matrix, $\mathbf{H}(\mathbf{x})$,[41] and the parameter t plays the role of x_{rc}. If $det(\mathbf{A}(\mathbf{x}))$ is positive definite along the whole NT curve joining two minimums of the PES, then this curve is a reaction path and it has the MEP category, because for this model curve both reaction path and MEP formulation coincide. In addition, in Ref. 16, it is shown that the second variation of the integral functional given in Eq.(17) is positive definite if the NT curve satisfies the inequality $\mathbf{g}(\mathbf{x})^T \mathbf{A}(\mathbf{x}) \mathbf{g}(\mathbf{x}) > 0$ which is noting more than the MEP requirement. Thus, for the NT model coincides the minimum variational condition with the MEP condition. However, if the NT has a turning point or a valley-ridged inflection point (VRI), then the minimum variational character is lost[16] and the reaction path and the MEP conditions are not satisfied. An NT curve can start at any point of the PES.

4 The Branching of a Reaction Path: Valley-Ridge-Inflection Points

The analysis of PESs remains an important basis for classifying and understanding the reasons of the mechanisms of chemical reactions as well as their dynamics. It is associated to the concept of the reaction path or to the definition of the minimum energy path on a PES. The chemical reaction may be composed by a number of elementary processes characterizing the mechanism of the reaction. Reaction path bifurcations are omnipresent on PESs; they happen at VRI points already on the PES of very small molecules like H_2O[43], H_2S, H_2Se, H_2CO[44], HCN[45,46], the ethyl cation[47], H_3CO, C_2H_5F[48], and many others. The importance of VRI points for the chemical reactivity is described in the reviews of Ess et al.[49,50] and Refs. 51,52. The intrinsic reaction coordinate (IRC) is a type of a reaction path which is widely used. However, in "skew", non-symmetric cases this curve usually does not meet a VRI point being nearby[53,54]. As mentioned in Section 3, there is a variety of types of curves that can be used as reaction path models, (a former ansatz was coordinate driving) which can be used in many cases to characterize the reaction path[16,34]. Sometimes the GE curves[17,29–31,55,56] also appear to form a suitable ansatz for such purposes. Certain NTs describe the valley or cirque structures of a PES, as well as their complements of ridges or cliffs (for the definition of such structures see Ref. 57. The structures are related to important chemical properties of the PES of the reaction under study[16,52]. The use of NTs opens the possibility to find and to study VRI points and, in succession, the bifurcation points or the branching points of reaction channels, because the reaction channel-branching is related to the existence of a special class of points of the PES, the VRI points[58,59]. A VRI point is that point in the configuration space where, orthogonally to the gradient, at least one main curvature of the PES becomes zero[60]. This definition implies that the gradient vector is orthogonal to an eigenvector of the Hessian matrix where its eigenvalue is zero. Thus, at least one PES-direction orthogonally to the gradient is flat. Usually, VRI points represent nonstationary points of the PES. Note that the VRI points are independent of the RP curve model used. They are related to the nature of the PES topography. Normally the VRI points are not related to the branching point of the reaction path curve except for NT curves[16,59]. So to say, a geometrical indicator of a VRI point is the bifurcation of a singular NT.

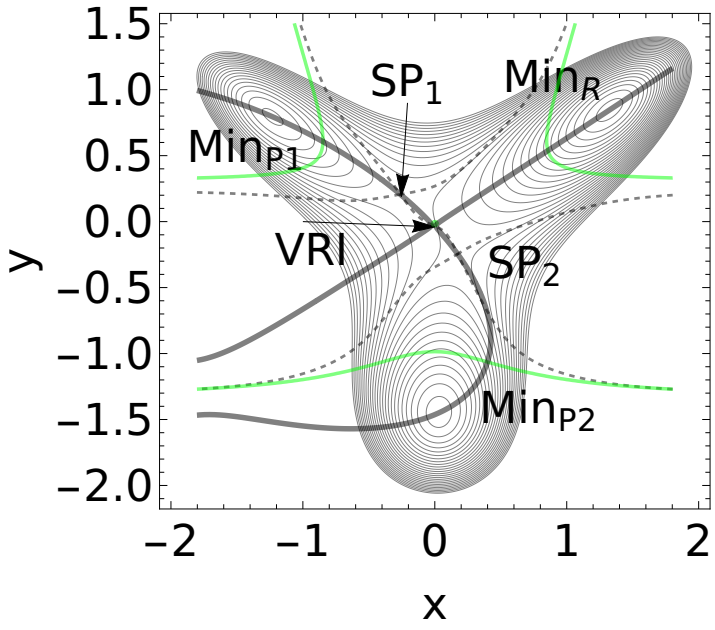

Figure 2: A modified BQC-PES with three minimums and two saddle points of index one. The green line corresponds to the condition $Det(\mathbf{H}(\mathbf{x})) = 0$. It indicates a change of the sign of an eigenvalue. The dashed line is the convexity border of the PES, namely, $Det[\mathbf{S}^T(\mathbf{x})\mathbf{H}(\mathbf{x})\mathbf{S}(\mathbf{x})] = 0$, where the matrix, $\mathbf{S}(\mathbf{x})$, is formed by the set of all orthonormalized vectors all of them are orthogonal to the $\mathbf{g}(\mathbf{x})$ vector. The fat line corresponds to a singular NT curve which crosses the VRI point.

The IRC curve is mathematically expressed through an autonomous system of differential equations for the tangent vector describing its evolution[13]. Its solution is unique; due to this fact no bifurcations can occur before reaching the next stationary point after the SP. No branching of PES valleys will be truly described or located by using the IRC curve as an RP type model[60,61]. It orthogonally traverses the family of levels, the equipotential energy surfaces[13]. Hirsch and Quapp[36] gave an example of a two-dimensional PES where the IRC is going over a skew ridge, however, it does not follow the valley ground nearby, which is here characterized by a GE. The IRC or any other SD curve does not take into account the curvature of the traversed contours in its evolution. In other words it does not give information on the valley floor or ridge character of its pathway. After a change of levels from convex to concave form the IRC curve ceases to be a valley pathway and is actually a merely RP. An early visualization of such an unstable minimum energy path was given by Mezey in Ref. 62, see also Ref. 63. As explained, the IRC curve traverses in its evolution a family of equipotential energy surfaces. At any point of an SD curve we can define a tangential plane to the equipotential energy surface orthogonally traversed

by the SD curve at the point, and the normal of the tangential plane is the gradient vector of the point. All direction vectors contained in the tangential plane are orthogonal to the gradient vector. If at least one of these direction vectors is connected with the curvature zero then we say that the steepest descent curve crosses aValley-Ridge transition (VRT) point. The curve leaves a valley and enters a ridge region of the PES or vice-versa. The VRT points are the border between valley- and ridge-regions. The concept of a VRT point is much more general than the VRI point concept. In fact a VRI point is a special case of a VRT point. In the general VRT situation, the gradient vector is not orthogonal to the set of eigenvectors of the Hessian matrix. This is the most general behavior. The zero curvature of the PES along the level line or equipotential energy surface at the VRT point comes from a suitable linear combination of the eigenvectors with their eigenvalues of the Hessian matrix. A manifold of points with these features exist on a PES. They are border points between quasi-convex valley regions and ridges. NTs there have a turning point. So to say, a turning point of an NT is the geometrical indicator of a VRT point, see Ref. 41. An example is given in Fig.2. This modified BQC-PES is near to the original reported in Ref. 17. Its equations is

$$V(x,y) = 1000/3(y^3 - 3yx^2) - 30(y + 3x) + 250((y + 0.7/4)^4 + x^4) \ . \quad (19)$$

5 Acknowledgements

Financial support from the Spanish Ministerio de Economía y Competitividad, Project CTQ2016-76423-P, is acknowledged.

References

[1] P. Brumer, M. Shapiro, Chem.Phys.Lett. **126**, 541 (1986)

[2] D.J. Tannor, R. Kosloff, S.A. Rice, J.Chem.Phys. **85**, 5805 (1986)

[3] E.D. Potter, J.L. Herek, S. Pedersen, Q. Liu, A.H. Zewail, Nature **355**, 66 (1992)

[4] A.H. Zewail, J. Phys. Chem. A **104**, 6660 (2000)

[5] A.P. Peirce, M.A. Dahleh, H. Rabitz, Phys. Rev. A: At., Mol., Opt. Phys. **37**, 4950 (1988)

[6] W. Zhu, H. Rabitz, J. Chem. Phys. **109**, 385 (1998)

[7] K. Rani, V. Rao, Bioprocess Eng. **21**, 77 (1999)

[8] J.M. Bofill, W. Quapp, Theor. Chem. Acc. **135**, 11 (2016)

[9] R. McWeeny, *Methods of Molecular Quantum Mechanics, 2nd ed.* (Academic Press, Oxford, 1992)

[10] R. Carbó, J.M. Riera, *A General SCF Theory.* Lecture Notes in Chemistry, Vol. 5 (Springer, New York, 1976)

[11] J.M. Bofill, H. Bono, J. Rubio, J. Comput. Chem. **19**, 368 (1998)

[12] D. Heidrich, W. Kliesch, W. Quapp, *Properties of Chemically Interesting Potential Energy Surfaces* (Springer, Berlin, Heidelberg, 1991)

[13] R. Crehuet, J.M. Bofill, J. Chem. Phys. **122**, 234105 (2005)

[14] A. Aguilar-Mogas, R. Crehuet, X. Giménez, J.M. Bofill, Mol. Phys. **105**, 2479 (2007)

[15] A. Aguilar-Mogas, X. Giménez, J.M. Bofill, J. Chem. Phys. **128**, 104102 (2008)

[16] J.M. Bofill, W. Quapp, J. Chem. Phys. **134**, 074101 (2011)

[17] J.M. Bofill, W. Quapp, M. Caballero, J. Chem. Theory Comput. **8**, 927 (2012)

[18] W. E, X. Zhou, Nonlinearity **24**, 1831 (2011)

[19] A.S. Householder, J. ACM **5**, 339 (1958)

[20] E. Zermelo, Z. Angew. Math. Mech. **11**, 114 (1931)

[21] C. Carathéodory, *Variationsrechnung und Partielle Differentialgleichungen erster Ordnung* (B. G. Teubner, Leipzig und Berlin, 1935)

[22] C. Carathéodory, Acta Math. **47**, 199 (1926)

[23] J.M. Bofill, W. Quapp, M. Caballero, Chem. Phys. Lett. **583**(0), 203 (2013)

[24] I.B. Vapnyarskii, Encyclopedia of Mathematics.

[25] L.S. Pontryagin, V.G. Boltyanski, R.V. Gamkrelidze., E.F. Mishechenko, *The Mathematical Theory of Optimal Processes* (John Wiley & Sons, NY, 1962)

[26] K. Fukui, J. Phys. Chem. **74**, 4161 (1970)

[27] W. Quapp, D. Heidrich, Theor. Chim. Acta **66**, 245 (1984)

[28] H. Hratchian, M.J. Frisch, H.B. Schlegel, J. Chem. Phys. **133**, 224101 (2010)

[29] M. Basilevsky, A. Shamov, Chemical Physics **60**, 347 (1981)

[30] D.K. Hoffmann, R.S. Nord, K. Ruedenberg, Theor. Chim. Acta **69**, 265 (1986)

[31] W. Quapp, Theoret. Chim. Acta **75**, 447 (1989)

[32] H.B. Schlegel, Theor. Chim. Acta **83**, 15 (1992)

[33] M.J. Rothman, L.L. Lohr, Chem. Phys. Lett. **70**, 405 (1980)

[34] W. Quapp, M. Hirsch, O. Imig, D. Heidrich, J. Comput. Chem. **19**, 1087 (1998)

[35] J.M. Anglada, E. Besalú, J.M. Bofill, R. Crehuet, J. Comput. Chem. **22**, 387 (2001)

[36] M. Hirsch, W. Quapp, Chem. Phys. Lett. **395**(1–3), 150 (2004)

[37] D.J. Rowe, A. Ryman, J. Math. Phys. **23**, 732 (1982)

[38] G.A. Bliss, *Lectures on the calculus of variations* (Chicago University Press, Chicago, IL, 1946)

[39] W. Quapp, J. Theoret. Comput. Chem. **2**, 385 (2003)

[40] W. Quapp, J. Chem. Soc., Faraday Trans. **90**, 1607 (1994)

[41] M. Hirsch, W. Quapp, J. Math. Chem. **36**, 307 (2004)

[42] F.H. Branin, IBM J. Res. Develop. **16**, 504 (1972)

[43] M. Hirsch, W. Quapp, D. Heidrich, Phys. Chem. Chem. Phys. **1**, 5291 (1999)

[44] W. Quapp, V. Melnikov, Phys. Chem. Chem. Phys. **3**, 2735 (2001)

[45] W. Quapp, B. Schmidt, Theor. Chem. Acc. **128**, 47 (2011)

[46] B. Schmidt, W. Quapp, Theor. Chem. Acc. **132**, 1305 (2012)

[47] W. Quapp, D. Heidrich, Journal of Molecular Structure: {THEOCHEM} **585**(1–3), 105 (2002). DOI 10.1016/S0166-1280(02)00037-4

[48] Y. Kumeda, T. Taketsugu, J. Chem. Phys. **113**, 477 (2000)

[49] D.H. Ess, S.E. Wheeler, R. Iafe, L. Xu, N. Çelebi-Ölçüm, K. Houk, Angew Chem Int Ed **47**, 7592 (2008)

[50] V. Bakken, D. Danovich, S. Shaik, H.B. Schlegel, J. Am. Chem. Soc. **123**, 130 (2001)

[51] W. Quapp, J. Bofill, A. Aguilar-Mogas, Theor. Chem. Acc. **129**, 803 (2011)

[52] W. Quapp, J.M. Bofill, J. Math. Chem. **50**(5), 2061 (2012)

[53] W. Quapp, J. Chem. Phys. **143**, 177101 (2015)

[54] Y. Harabuchi, Y. Ono, S. Maeda, T. Taketsugu, J. Chem. Phys. **143**, 014301 (2015)

[55] J.Q. Sun, K. Ruedenberg, J. Chem. Phys. **98**, 9707 (1993)

[56] K. Bondensgård, F. Jensen, J. Chem. Phys. **104**, 8025 (1996)

[57] G.J. Atchity, S.S. Xantheas, K. Ruedenberg, J. Chem. Phys. **95**, 1862 (1991)

[58] P. Valtazanos, K. Ruedenberg, Theor. Chim. Acta **69**, 281 (1986)

[59] W. Quapp, M. Hirsch, D. Heidrich, Theor. Chem. Acc. **100**(5/6), 285 (1998)

[60] W. Quapp, J. Mol. Struct. (THEOCHEM) **695-696**, 95 (2004)

[61] W. Quapp, M. Hirsch, D. Heidrich, Theor. Chem. Acc. **112**, 40 (2004)

[62] P. Mezey, *Potential Energy Hypersurfaces* (Elsevier, Amsterdam, 1987)

[63] P. Mezey, Theor. Chem. Acc. **54**, 95 (1980)