

Exploring Potential-Energy Surfaces with Gentlest Ascent Dynamics in Combination with the Shrinking Dimer Method and Newtonian Dynamics[±]

Guillem Albareda,^{1,4} Josep Maria Bofill,^{*2,4} Ibério de P.R. Moreira,^{*1,4} Wolfgang Quapp,³ Jaime Rubio-Martínez^{1,4}

¹*Departament de Ciència dels Materials i Química Física, Secció de Química Física, Universitat de Barcelona, Martí i Franquès, 1, 08028 Barcelona, Spain.*

²*Departament de Química Inorgànica i Orgànica, Secció de Química Orgànica, Universitat de Barcelona, Martí i Franquès, 1, 08028 Barcelona, Spain.*

³*Department of Mathematics, University Leipzig, PF 100920, D-04009 Leipzig, Germany.*

⁴*Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franquès, 1, 08028 Barcelona, Spain.*

Abstract

We revisit the so-called Gentlest Ascent Dynamics reaction path model for finding saddle points of any index in multidimensional potential-energy surfaces. The variational nature of the method is analyzed in detail and an algorithm for the integration of its equations of motion is proposed based on the optimization-based shrinking dimer method. By means of three different two-dimensional model potential-energy surfaces, we argue that the combination of the proposed method with Newtonian (dissipative) dynamics could lead to a practical scheme for the exhaustive exploration of potential-energy landscapes.

[±] *Dedicated to the memory of Professor Claudio María Zicovich-Wilson.*

I. Introduction

A crucial achievement in the development of models to understand the chemical reaction mechanisms was the introduction of the following two concepts, namely, the Born-Oppenheimer potential-energy surface (PES) and the reaction path (RP). They provide a strategy to describe and analyze a molecular system evolution from reactants to products in geometrical terms [1, 2]. In the context of molecular conformational rearrangements, minima and saddle points (or transition states) on the PES, and the paths joining these points are the most important features. In the present context, minima correspond to low energy molecular conformations and saddle points correspond to transition states associated with conformational inter-conversions. Saddle points provide the lowest energy path across a PES barrier. Therefore, the main goal and purpose of a conformational study consists into finding the set of RPs that correspond to a network of conformational inter-conversions.

The so-called eigenvector (or mode) following (EF) method has been used since many years to explore molecular conformations, see e.g. [3]. This method allows to find saddle points in multidimensional PES by starting from a given minimum (either local or global). The EF method applied to conformational molecular studies was first proposed by Crippen and Scheraga [4], who called the method gentlest ascent. The authors describe an algorithm for passing from a minimum position in a many-dimensional PES to a neighboring minimum via an intervening saddle point. An alternative application of the EF method is due to Nakamura and co-workers [5]. The authors use the mode following in combination with the diffusion equation method to find the global minimum energy of a small peptide. An alternative algorithm to explore PESs that uses the lowest eigenmode of the Hessian matrix is the activation-relaxation method (ART) proposed by Barkema and Mousseau [6]. Recent modifications of ART have significantly enhanced its convergence properties [7, 8]. These improvements in the equations of motion involve the calculation of the minimum eigenvalue and its eigenvector. It has been shown that the improved ART converges quadratically to the saddle point along the eigenmode of negative curvature while it converges linearly in perpendicular directions. More recently it has been proposed the gentlest ascent dynamics (GAD) [9-14]. Also based on the eigenmode philosophy, it consists of a set of dynamic equations whose solutions converge to saddle points. The stability of these equations and their convergence close to stationary points on the PES was already analyzed [10].

In all the methods summarized above, evaluating the minimum eigenmode of the corresponding Hessian matrix and taking it as the initial exploring direction is crucial in order to converge to an index-1 saddle point. However, the evaluation and diagonalization of the Hessian matrix is feasible for small molecular systems, but it becomes computationally very expensive for molecules with more than ~ 100 atoms for self-consistent-field type calculations or for molecules with more than a very few degrees of freedom for highly correlated wavefunction calculations. For this reason, many techniques have been proposed to evaluate the minimum eigenmode of the Hessian in an optimal and inexpensive way. An example is the shrinking dimer method (SD) proposed by Henkelman and Jonsson [15], which allows to find the ascent direction that characterizes the final reaction path without constructing the full Hessian matrix. The method only makes use of first derivatives of the PES and it is, therefore, applicable in situations where second derivatives are computationally inaccessible or very expensive. A dimer consists of two points in the configuration space separated by a

small (parametric) distance. The orientation of the dimer is determined by minimizing the so-called dimer energy [15]. The process of obtaining the minimum eigenmode of the Hessian is then approximated by first selecting a suitable plane spanned by the rotational force acting on the dimer and subsequently selecting the dimer orientation. The dimer is then rotated in this plane to obtain the optimal direction corresponding to the minimum energy of the dimer [15].

In this work we revisit the GAD method to find saddle points in multi-dimensional PESs and propose an implementation of the method by combining it with an optimization-based version of the SD method due to L. Zhang and co-workers [20]. By means of three simple two-dimensional model PESs, we show that the combination of the GAD-SD method with Newton-like dynamics leads to a novel computational scheme for the exploration of PESs. The article is organized as follows. In Section II we discuss the variational nature of GAD equations (the mathematical proof is provided in Appendix A). In Section III an algorithm is described to integrate the GAD equations of motion with the help of the SD method for the case of stationary points of index $n = 1$ (the extension of this algorithm to any dimension is reported in Appendix B). By means of three examples, in Section IV we will present the GAD-SD algorithm in combination with dissipative Newton dynamics as a novel “metadynamics” for PESs.

II. The Variational Nature of the Gentlest Ascent Dynamics Reaction Path Model

Consider a molecular system consisting of N nuclear degrees of freedom collectively denoted by the position vector \mathbf{x} and the PES $V(\mathbf{x})$. The gradient vector at a point \mathbf{x} is $\mathbf{g}(\mathbf{x}) = \nabla_{\mathbf{x}}V(\mathbf{x})$ and the Hessian matrix is given by $\mathbf{H}(\mathbf{x}) = \nabla_{\mathbf{x}}\nabla_{\mathbf{x}}^TV(\mathbf{x})$, where the superscript T means transposition. The equations that describe the GAD method are [9-14]:

$$\frac{d\mathbf{x}}{dt} = -[\mathbf{I} - 2\mathbf{W}\mathbf{W}^T]\mathbf{g}(\mathbf{x}) \quad (1)$$

$$\frac{d\mathbf{W}}{dt} = -[\mathbf{I} - \mathbf{W}\mathbf{W}^T]\mathbf{H}(\mathbf{x})\mathbf{W} \quad (2)$$

where \mathbf{I} is the unit matrix of dimension N , and $\mathbf{W} = [\mathbf{w}^{(1)} | \dots | \mathbf{w}^{(n)}]$, being $n \leq N$. The matrix \mathbf{W} satisfies $\mathbf{W}^T\mathbf{W} = \mathbf{I}_n$, where \mathbf{I}_n is the unit matrix of dimension n . In Equation (1) the matrix $[\mathbf{I} - 2\mathbf{W}\mathbf{W}^T]$ is a mirror transformation at the mirror hyperplane defined by the set of vectors $\{\mathbf{w}^{(i)}\}_{i=1}^n$. This matrix reverses the $-\mathbf{g}(\mathbf{x})$ vector in the hyperplane. Equation (2) defines the evolution of the set of orthonormalized directions, $\{\mathbf{w}^{(i)}\}_{i=1}^n$. In the case that $n = N$, Equation (1) coincides with the steepest ascent expression, whereas for the situation that $\mathbf{W}^T\mathbf{g}(\mathbf{x}) = \mathbf{0}_n$, Equation (1) coincides with the steepest descent. The curve described by Equations (1) and (2) is attracted at saddle points of index n [10].

The GAD model is variational in nature [14]. More precisely, the path described by the GAD equations to find transition states on a PES is an example of a quickest “nautical” path for a given stationary wind or current, i.e., the so-called Zermelo navigation variational problem [16-17]. Since the Zermelo navigation problem is an example of optimization control theory, GAD satisfies the Poytriagin Maximum Principle [18]. In

reference [14] it is proved that Equations (1) and (2) are the canonical equations of the Hamiltonian

$$2H(\mathbf{x}, \mathbf{y}) = 4(\mathbf{g}^T(\mathbf{x})\mathbf{W}\mathbf{W}^T\mathbf{g}(\mathbf{x}))\mathbf{y}^T\mathbf{y} - (1 + \mathbf{g}^T(\mathbf{x})\mathbf{y})^2 = 0 \quad (3)$$

where \mathbf{y} is the conjugate canonical coordinate of \mathbf{x} and directly related with the set of directions $\{\mathbf{w}^{(i)}\}_{i=1}^n$. Equation (3) is the Gentlest Ascent Hamiltonian. For this reason we re-name it here the Gentlest Ascent Zermelo model. In Appendix A we provide a variational proof and a discussion on the minimum character for the case of index $n = 1$.

III. Integration of the Gentlest Ascent Dynamics Zermelo Equations in combination with the Shrinking Dimer method

We propose the optimization-based shrinking dimer (SD) method by using the step-size selection of the gradient due to Barzilai and Borwein [19, 20] to solve the systems of Equations (1) and (2). Briefly, we apply the SD method with optimized steps for the case of $n = 1$ such that the \mathbf{W} matrix is reduced to the vector $\mathbf{w}^{(1)} = \mathbf{w}$. The subscripts indicate the iteration number.

Let \mathbf{x}_i and \mathbf{x}_{i+1} two points of the PES separated by a distance $d_i = ((\mathbf{x}_{i+1} - \mathbf{x}_i)^T(\mathbf{x}_{i+1} - \mathbf{x}_i))^{1/2}$. These two points define the dimer and an inner point of the dimer is given by $\mathbf{x}_{c_i} = (1 - p_i)\mathbf{x}_i + p_i\mathbf{x}_{i+1}$ where $0 < p_i < 1$. The dimer orientation is given by the normalized control vector \mathbf{w}_i , $\mathbf{x}_{i+1} - \mathbf{x}_i = d_i\mathbf{w}_i$. In this way $\mathbf{x}_i = \mathbf{x}_{c_i} - p_i d_i\mathbf{w}_i$ and $\mathbf{x}_{i+1} = \mathbf{x}_{c_i} + (1-p_i)d_i\mathbf{w}_i$. The gradient at the central point of the dimer is (see Appendix C for the proof)

$$\mathbf{g}(\mathbf{x}_{c_i}) = (1 - p_i)\mathbf{g}(\mathbf{x}_i) + p_i\mathbf{g}(\mathbf{x}_{i+1}). \quad (4)$$

A direct way for solving Equation (2) for $n = 1$ is to use the approximate $(\mathbf{g}(\mathbf{x}_{i+1}) - \mathbf{g}(\mathbf{x}_i)) = \mathbf{H}(\mathbf{x}_{c_i})\mathbf{w}_i d_i + O(d_i^2)$ and

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \Delta t_i [\mathbf{I} - \mathbf{w}_i\mathbf{w}_i^T] (\mathbf{g}(\mathbf{x}_{i+1}) - \mathbf{g}(\mathbf{x}_i)) / d_i. \quad (5)$$

For the dimer translation we solve Equation (1) for $n = 1$ using the approximation

$$\mathbf{x}_{c_{i+1}} = \mathbf{x}_{c_i} - \Delta t_i [\mathbf{I} - 2\mathbf{w}_i\mathbf{w}_i^T] \mathbf{g}(\mathbf{x}_{c_i}). \quad (6)$$

where $\mathbf{g}(\mathbf{x}_{c_i})$ is given in Equation (4). In the same philosophy as the shrinking dimer method [21], following the dimer rotation and translation steps, we shrink the dimer length d_i and force it to approach zero in order to guarantee the convergence, in other words, $d_i \rightarrow 0$ as $i \rightarrow \infty$. Now we write an algorithm by considering Equations (5) and (6) with the above condition on d_i with the purpose to minimize the functional of Equation (A.2).

We collect the vectors \mathbf{x}_{c_i} and \mathbf{w}_i into the vector \mathbf{h}_i , namely, $\mathbf{h}_i^T = (\mathbf{x}_{c_i}^T, \mathbf{w}_i^T)$, and the so-called residues into the vector \mathbf{r}_i

$$\mathbf{r}_i = \begin{pmatrix} [\mathbf{I} - 2\mathbf{w}_i\mathbf{w}_i^T]\mathbf{g}(\mathbf{x}_{ci}) \\ [\mathbf{I} - \mathbf{w}_i\mathbf{w}_i^T](\mathbf{g}(\mathbf{x}_{i+1}) - \mathbf{g}(\mathbf{x}_i)) / d_i \end{pmatrix}. \quad (7)$$

With these definitions the set of Equations (5) and (6) can be written in a compact form as

$$\Delta\mathbf{h}_i = -\Delta t_i \mathbf{r}_i \quad (8)$$

where $\Delta\mathbf{h}_i = \mathbf{h}_{i+1} - \mathbf{h}_i$.

While not strictly necessary, it is often convenient to introduce now the quasi-Newton property, $\Delta\mathbf{h}_{i-1} = -\Delta t_i \Delta\mathbf{r}_{i-1}$, where $\Delta\mathbf{r}_{i-1} = \mathbf{r}_i - \mathbf{r}_{i-1}$ and look for Δt_i that minimizes $(\Delta\mathbf{h}_{i-1} + \Delta t_i \Delta\mathbf{r}_{i-1})^T (\Delta\mathbf{h}_{i-1} + \Delta t_i \Delta\mathbf{r}_{i-1})$, resulting in

$$\Delta t_i = -\frac{\Delta\mathbf{h}_{i-1}^T \Delta\mathbf{r}_{i-1}}{\Delta\mathbf{r}_{i-1}^T \Delta\mathbf{r}_{i-1}}. \quad (9)$$

It yields the shrinking condition $d_{i+1} = d_i / (1 + \Delta t_0)$. Otherwise the dimer distance is updated as $d_i = ((\mathbf{x}_{i+1} - \mathbf{x}_i)^T (\mathbf{x}_{i+1} - \mathbf{x}_i))^{1/2}$.

Combining Equations (8) and (9) and the shrinking condition $d_i \rightarrow 0$ as $i \rightarrow \infty$ we can already introduce the following algorithm:

- 1) given the initial conditions $\mathbf{h}_i, \mathbf{r}_i, d_i, \Delta t_i$ ($\mathbf{h}_0, \mathbf{r}_0, d_0, \Delta t_0$ at $t = 0$)
- 2) apply Eq. (8) (also (9) if convenient) to update $\mathbf{h}_i, \mathbf{r}_i, d_i$
- 3) the iteration stops when $\mathbf{g}^T(\mathbf{x}_{ci+1}) \mathbf{g}(\mathbf{x}_{ci+1}) \leq th$

where th (abbreviation of threshold) is a parameter defined in Table II. In all our simulations we kept the time-step constant. The above numerical algorithm does not preserve the unit length of \mathbf{w}_i . For this reason we use a simple renormalization to ensure, $\mathbf{w}_i^T \mathbf{w}_i = 1$, at each iteration step. The extension of this algorithm to any control dimension n is reported in Appendix B.

IV. Combining the GAD-SD method with dissipative Newton dynamics: towards a novel approach for exploring PESs

In this section we provide examples of the previous algorithm to integrate the GAD-SD equations of motion in combination with dissipative classical dynamics. Our goal is to show that such a combined method could lead to a novel technique to explore multidimensional PESs. The idea is simple, we already know that the solution of the GAD-SD equations of motion represent the fastest (variational) way to reach a saddle point starting from a given position in the multiconfiguration space. Hence, if we are able to concatenate this technique with a dynamics that brings us from a saddle point to a minimum then we could already envision a numerical technique for the exploration of PESs.

To start with, we look at the behavior of the GAD-SD method, Equations (8)-(9), in combination with a dissipative Newton equation of the type,

$$\frac{d^2\mathbf{x}}{dt^2} = -\mathbf{g}(\mathbf{x}) - \gamma \frac{d\mathbf{x}}{dt} \quad (10)$$

where $\gamma > 0$ is a friction coefficient (see Table II). To give physical meaning to the friction force in the context of energy dissipation, i.e., the friction removes energy from the system until the trajectory reaches the minimum of the PES. In the trajectory framework, adding momentum-dependent friction to the equations of motion is a straightforward modification. Notice that the force of friction is often taken for processes in condensed phase [38] as linear in velocity of a particle.

We carry out calculations on three different PESs, viz., a model PES for the study of the double-proton transfer in porphine [35, 36], the so-called Wolfe-Quapp PES [22, 23], and the Rastrigin surface [37]. The mathematical expressions for these two PESs can be found in Table I. Both the porphine and the Wolfe-Quapp PESs contain quartic and quadratic terms in x and y and the coupling between the two coordinates is bilinear. The Rastrigin PES is nonlinear by trigonometric terms and its coupling is quadratic.

The general algorithm works as follows. The dynamics always starts with the GAD-SD equations of motion (8)-(9), and the initial conditions for a given trajectory are chosen randomly (both for the initial position and the control vector). As soon as the GAD-SD trajectory reaches a saddle point (which we identify by defining the threshold condition $|\mathbf{g}(\mathbf{x})|^2 < th$. See Table II), the algorithm switches from GAD-SD to dissipative Newtonian dynamics. After a small random boost ($\Delta\mathbf{x}$) on the position of the trajectory, the equations of motion (8)-(9) are substituted by Eq. (10), and the trajectory starts to fall into a neighboring (local or global) minimum. As soon as a minimum is reached (which we identify by checking the two conditions $\det(\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}^T V(\mathbf{x})) > 0$ and $|\mathbf{g}(\mathbf{x})|^2 < th$), the algorithm switches back from Newtonian to GAD-SD dynamics. A random boost is applied again on the trajectory and after an initial random control vector is chosen, the algorithm continues until all minima and saddle points have been sampled.

In Figs. (1) and (2) we show four randomly initialized trajectories for the porphine and the Wolfe-Quapp PESs respectively (the reader can find the numerical parameters used for each PES in Table II). Irrespectively of the initial random position and control vectors (the $\mathbf{h}_0 = (x_0, w_0)^T$ selected for each trajectory and PES can be found in Table III), the trajectories depicted in Figs. (1) and (2) all repeatedly pass over a transition state and fall down into a neighboring minimum until the entire PES has been explored. Regarding the parameterization of the algorithm, notice that the value of γ is chosen by trial with the goal of balancing the propagation time (more time is required if γ is small) and numerical stability (shorter time-steps are required if γ is large). Furthermore, the choice of the threshold parameter th strongly depends on the topography of the underlying PES, i.e., it is the steepness of the maxima that defines its value. Hence, for such a preliminary algorithm, PESs combining both steep and flat maxima would suffer from an ill-defined threshold parameter.

Notice that the GAD-SD method, as described above, does not require the computation of the Hessian matrix at any time, and only the conditions for reaching a minimum require the evaluation of second order derivatives. This would be especially welcome in

the context of highly correlated wavefunctions calculations (beyond self-consistent-field methods) where second order derivatives are much more expensive than first order derivatives. In these cases, optimization and path following will benefit from a gradient only algorithm. Anyhow we can say that the combination of the GAD-SD method with dissipative Newtonian dynamics seems to be a good starting point to develop a new type of “metadynamics” approach [24-26].

The efficiency and numerical stability of the method for larger (multidimensional) systems [27, 28] will be reported elsewhere. Nonetheless, we here want to give some clues on the viability of this extension to larger systems by testing the method on the so-called Rastrigin function [37]. The Rastrigin function contains only up to quadratic terms in x and y , but the coupling between the two is quadratic. Furthermore, the Rastrigin function is a non-convex function used as a performance test problem for optimization algorithms. It is a typical example of non-linear multi-modal function. Finding the minima and saddle points of this function is a fairly difficult problem due to its large search space and its large number of local minima. Therefore, in order to improve the efficiency of our algorithm in the presence of multiple reaction paths, we reinitialize our trajectory every time it finds an already known minimum or saddle point. In this way we avoid exploring paths that have been already scrutinized. In Fig. (3) we show that, after a very few trajectory resets, the entire surface has been sampled. Notice that while in our algorithm the different trajectories are evolved in series, one could also envision an equivalent algorithm where a swarm of trajectories are launched all at once in parallel. Such a strategy could be even more appropriate for real systems with a larger number of degrees of freedom.

Table 1. Mathematical expressions for the three PESs used in this work. The three parameters of the porphine PES are: $U_0 = 0.01783$, $G = 0.063$, $A = 1.251$. The parameters for the Wolfe-Quapp PES are: $A = -2$, $B = -4$, $C = 0.3$, and $D = 0.1$. Finally, the two parameters of the Rastrigin surface are $A = 2$, and $N = 0.1$.

	$V(x,y)$
Porphine	$U_0 / A^4 [(x^2 - A^2)^2 + (y^2 - A^2)^2 - 4G \cdot A^2 xy] + 2G(2 + G)U_0$
Wolfe-Quapp	$x^4 + y^4 + Ax^2 + By^2 + xy + Cx + Dy$
Rastrigin	$A \cdot N + x^2 - A \cos(2\pi x) + y^2 - A \cos(2\pi y) + (xy)^2$

Table 2. Description of the five parameters required in the present simulations for the three different PESs.

	Porphine	Wolfe-Quapp	Rastrigin
th	$1 \cdot 10^{-6}$	$3 \cdot 10^{-3}$	$5 \cdot 10^{-1}$
γ	2	4	5
Δx	$4 \cdot 10^{-2}$	$8 \cdot 10^{-3}$	$4 \cdot 10^{-2}$
$\Delta t = \Delta t_0$	$3 \cdot 10^{-2}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-2}$
d_0	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$

Table 3. Randomly chosen initial positions and control vectors for the trajectories depicted in Figs. (1) – (3).

	Porphine	Wolfe-Quapp	Rastrigin
<i>Traj. 1</i>	$x_0 = (-0.7, 0.9)$ $w_0 = (-1.0, 0.9)$	$x_0 = (0.1, 0.3)$ $w_0 = (0.6, -0.9)$	$x_0 = (0.4, -1.3)$ $w_0 = (0.2, 0.5)$
<i>Traj. 2</i>	$x_0 = (0.7, 0.0)$ $w_0 = (0.2, -0.6)$	$x_0 = (1.4, -1.3)$ $w_0 = (-0.6, -0.7)$	
<i>Traj. 3</i>	$x_0 = (-0.5, 0.6)$ $w_0 = (-0.8, 0.5)$	$x_0 = (-0.4, -1.0)$ $w_0 = (0.6, -0.1)$	
<i>Traj. 4</i>	$x_0 = (-0.1, 0.2)$ $w_0 = (-0.1, 0.5)$	$x_0 = (0.0, 0.3)$ $w_0 = (0.1, 0.2)$	

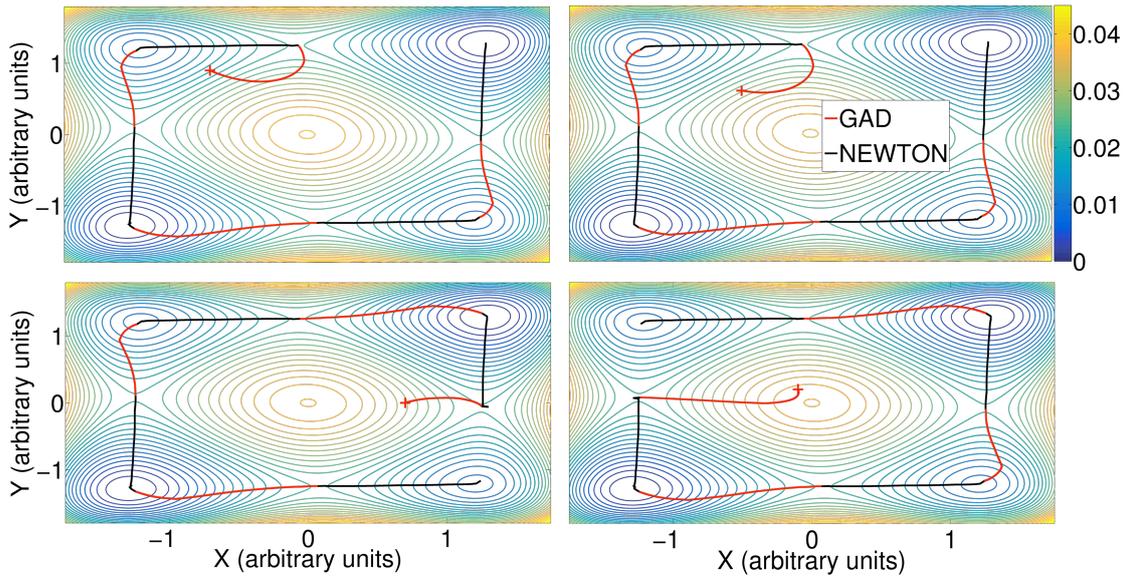


Figure 1. Four randomly initialized trajectories for the combined model GAD-SD-Newtonian dynamics on the model porphine. Starting from four different positions and control vectors (see Table III), the GAD-SD trajectories rapidly climb to a saddle point and, after switching to classical dissipative dynamics, they fall into a neighboring valley of the PES. This process is repeated as many times as required to sample the entire PES.

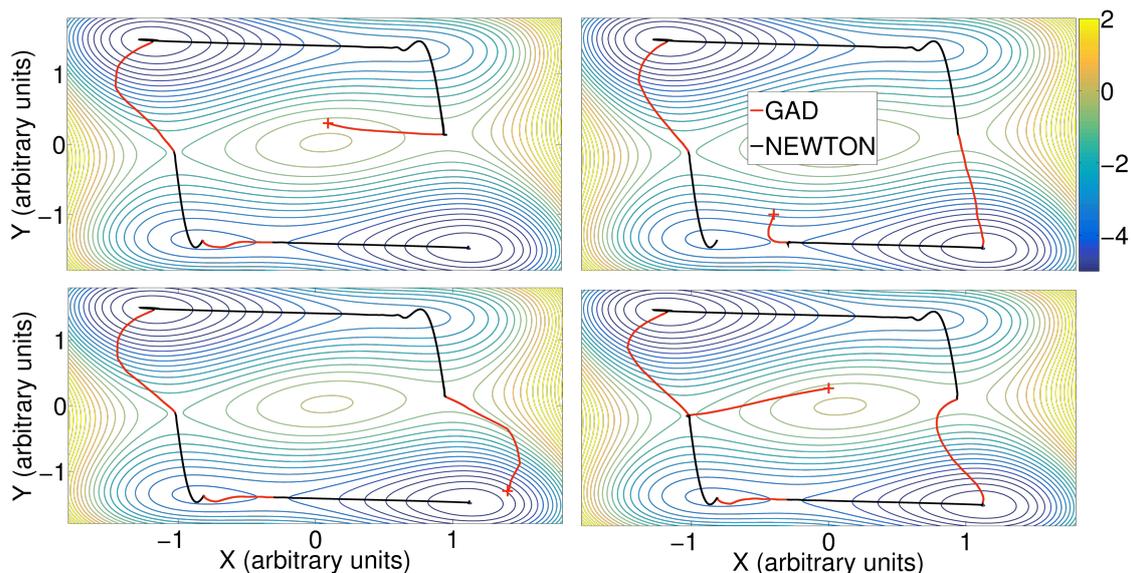


Figure 2. The same as in Fig. 1 but for the Wolfe-Quapp PES. Starting from four different random positions and control vectors (see Table III), all GAD-SD trajectories combined with classical dissipative dynamics are successfully exploring the PES.

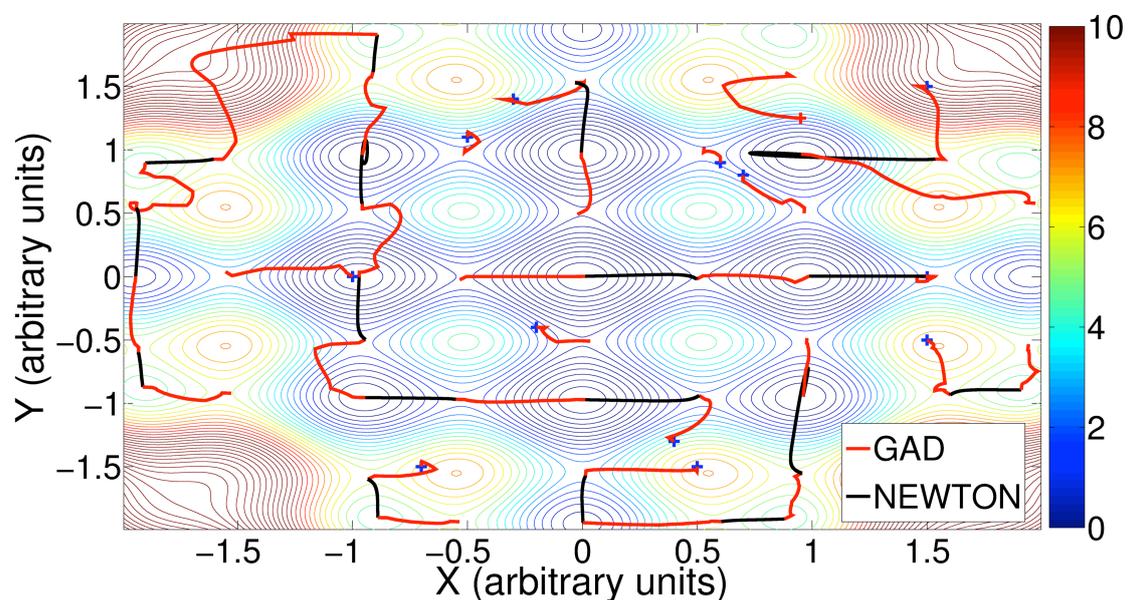


Figure 3. Due to the large number of possible reaction paths co-existing on the Rastrigin PES, the GAD-SD-Newtonian trajectory is restarted as soon as an already known saddle point or minimum is reached. In this way a given path along the PES is sampled only once and the efficiency of the algorithm is considerably improved.

VII. Conclusions

We have presented an algorithm to integrate the GAD equations of motion for finding saddle-points of index n based on an optimization-based SD method. More specifically, the Barzilai and Borwein gradient method [19] is applied, and according to the results of Zhang et al. [20], this makes the resulting GAD-SD equations of motion stable. By combining the GAD-SD method with classical dynamics with friction, we report a promising algorithm to exhaustively exploring reaction paths in multi-dimensional PESs. The combined method has been first tested on three simple two-dimensional model PES. Independently on the initial position and control vector of the trajectories, all were able to find all saddle points and minima in a trivial manner.

Our results are not conclusive regarding the scalability of the method for a large number of degrees of freedom. However, we gave some clues on the viability of the method for real systems. By choosing the Rastrigin surface, we stressed our approach in a PES where multiple minima and saddle points (and hence multiple reaction paths) coexist, which is closer to a real molecule scenario. By reinitializing the trajectories one can avoid the repeated exploration of already known regions of the surface. The algorithm can be equivalently seen as launching a swarm of trajectories in parallel.

Let us emphasize that the GAD-SD method does not require the computation of the Hessian matrix at any time, and only the conditions for reaching a minimum require the evaluation of second order derivatives. This would be especially welcome in the context of highly correlated wavefunctions calculations (beyond self-consistent-field methods) where second order derivatives are computationally very expensive. In these cases, optimization and path following will benefit from a gradient only algorithm. Further studies regarding the extension of the method for larger systems or its efficiency in comparison to well-established methods for the exploration of multidimensional PESs will be reported soon elsewhere.

Acknowledgements

Financial support from the Spanish Ministerio de Economía y Competitividad, Project CTQ2016-76423-P, is acknowledged. G.A. acknowledges financial support from the Beatriu de Pinós program through the Project 2014-BP-B-00244.

Appendix A: Proof of the Variational Character of Gentlest Ascent-Zermelo Model for index $n = 1$.

The GAD-Zermelo problem can be formulated as a Mayer-Bolza problem of the Calculus of Variations related with the solution of a problem of Optimal Control Theory [14, 18]. We present the proof for the case of index $n = 1$, thus the \mathbf{W} matrix is reduced to a vector, $\mathbf{w}^{(1)} = \mathbf{w}$, the generalization is straightforward. This problem can be

formulated as follows: determine the vector function $\mathbf{x}(t)$ of dimension N , satisfying the equation

$$d\mathbf{x} / dt = -[\mathbf{I} - 2\mathbf{w}\mathbf{w}^T] \mathbf{g}(\mathbf{x}) \quad (\text{A.1})$$

where, $\mathbf{g}(\mathbf{x})$, is the gradient vector of the function $V(\mathbf{x})$, \mathbf{I} is the identity matrix, and the initial condition $\mathbf{x}_0 = \mathbf{x}(t_0)$, and determine the control vector function, $\mathbf{w}(t)$, of dimension N , restricted to the normalization relation, $\mathbf{w}^T \mathbf{w} = 1$, in such a way that the functional

$$J[\mathbf{x}(t, \mathbf{w}(t))] = t - t_0 \quad (\text{A.2})$$

assumes an extremal value with the boundary condition

$$\nabla_{\mathbf{x}} V(\mathbf{x}(t_f)) = \mathbf{g}(\mathbf{x}(t_f)) = 0 \quad (\text{A.3})$$

being satisfied at $t = t_f$.

If we take t_0 fixed then we can write

$$t - t_0 = \int_{t_0}^t dt = \int_{\tau_0}^{\tau} \frac{dt}{d\tau} d\tau = \int_{\tau_0}^{\tau} (\nabla_{\mathbf{x}} J)^T \left(\frac{d\mathbf{x}}{d\tau} \right) d\tau = \int_{\tau_0}^{\tau} F(\mathbf{x}, d\mathbf{x} / d\tau) d\tau \quad (\text{A.4})$$

where the last equality is satisfied if the last integral is evaluated through the extremal curve, Equation (A.1). According to the above equalities we have the relation, $F(\mathbf{x}, d\mathbf{x} / d\tau) = dt / d\tau$, through the extremal curve. From this result and in order to determine the optimal curve we must establish $F(\mathbf{x}, d\mathbf{x} / d\tau)$ as the basic functional of a variational problem. The functional $F(\mathbf{x}, d\mathbf{x} / d\tau)$ is assumed to be homogenous of degree one with respect to the argument $d\mathbf{x} / d\tau$. However, the tangent of the curve given in Equation (A.1) satisfy the equation

$$\frac{d\mathbf{x}}{dt} = \frac{d\mathbf{x}}{d\tau} \frac{d\tau}{dt} = \frac{d\mathbf{x}}{d\tau} \frac{1}{F(\mathbf{x}, d\mathbf{x} / d\tau)} = -[\mathbf{I} - 2\mathbf{w}\mathbf{w}^T] \mathbf{g}(\mathbf{x}) \quad (\text{A.5})$$

and from the last two equalities we write

$$\frac{1}{F(\mathbf{x}, d\mathbf{x} / d\tau)} \frac{d\mathbf{x}}{d\tau} + \mathbf{g}(\mathbf{x}) = (2\mathbf{w}^T \mathbf{g}(\mathbf{x})) \mathbf{w}. \quad (\text{A.6})$$

Now we rewrite Equation (A.6)

$$(\mathbf{z} + \mathbf{g}(\mathbf{x}))^T (\mathbf{z} + \mathbf{g}(\mathbf{x})) - (2\mathbf{w}^T \mathbf{g}(\mathbf{x}))^2 = 0 \quad (\text{A.7})$$

where $\mathbf{z} = (1/F(\mathbf{x}, d\mathbf{x} / d\tau)) d\mathbf{x} / d\tau$. The Equation (A.7) defines a hypersurface in the space of \mathbf{z} which depends on the parameters \mathbf{x} and \mathbf{w} and which is the indicatrix of the present variational problem at the point (\mathbf{x}, \mathbf{w}) . This geometrical interpretation of the

Calculus of Variations is very convenient. It allows the function $F(\mathbf{x}, d\mathbf{x} / d\tau)$ or a Hamiltonian function of the present problem to be found. In particular, due to the form of Equation (A.7) it allows us to calculate the function $F(\mathbf{x}, d\mathbf{x} / d\tau)$ as a positive root if such a root exists. However, we can obtain a Hamiltonian function H from the indicatrix by calculating the relation between the variables $(\mathbf{x}, \mathbf{w}, \mathbf{y})$, being $\mathbf{y} = \nabla_{d\mathbf{x}/d\tau} F(\mathbf{x}, \mathbf{z})$, which arises first by multiplication the Equation (A.7), by $F^2(\mathbf{x}, d\mathbf{x} / d\tau)$, and second by differentiation with respect to $d\mathbf{x} / d\tau$ and finally substituting in the resulting expression Equation (A.6) obtaining

$$(\mathbf{I} + \mathbf{y}\mathbf{g}^T(\mathbf{x}))\mathbf{w} = (2\mathbf{w}^T\mathbf{g}(\mathbf{x}))\mathbf{y}. \quad (\text{A.8})$$

In this way we have eliminated the \mathbf{z} vector. In order to simplify Equation (A.8) we now set the eigenvalue equation,

$$(\mathbf{I} + \mathbf{y}\mathbf{g}^T(\mathbf{x}))\mathbf{y} = (1 + \mathbf{y}^T\mathbf{g}(\mathbf{x}))\mathbf{y} = \omega\mathbf{y} \quad (\text{A.9})$$

then multiplying by ω Equation (A.8) we have the following equation

$$\omega\mathbf{w} = (2\mathbf{w}^T\mathbf{g}(\mathbf{x}))\mathbf{y}. \quad (\text{A.10})$$

If we multiply Equation (A.10) from the left by $\mathbf{g}^T(\mathbf{x})$ and if we take the value of ω , given in Equation (A.9), then we obtain that $\mathbf{y}^T\mathbf{g}(\mathbf{x}) = 1$ and $\omega = 2$. Moreover, it follows from Equation (A.10) that

$$2H(\mathbf{x}, \mathbf{y}) = (2\mathbf{w}^T\mathbf{g}(\mathbf{x}))^2 \mathbf{y}^T\mathbf{y} - \omega^2 = (2\mathbf{w}^T\mathbf{g}(\mathbf{x}))^2 \mathbf{y}^T\mathbf{y} - (1 + \mathbf{y}^T\mathbf{g}(\mathbf{x}))^2 = 0 \quad (\text{A.11})$$

which is the Hamiltonian function being a quadratic function of the \mathbf{y} vector. This surface, $H(\mathbf{x}, \mathbf{y}) = 0$, in the present case is also called figuratrix. The indicatrix and figuratrix are reciprocal polars with respect to a sphere of radius equal one. This construction is due to the inapplicability of Legendre transformation to obtain the Hamiltonian. The homogeneous character of degree one of the functional $F(\mathbf{x}, d\mathbf{x} / d\tau)$ with respect to the tangent avoids its applicability [17]. We can derive the canonical equations for the extremal curves

$$\frac{d\mathbf{x}}{d\tau} = \lambda \nabla_{\mathbf{y}} H(\mathbf{x}, \mathbf{y}) = \lambda H_{\mathbf{y}}(\mathbf{x}, \mathbf{y}) = \lambda \left[(2\mathbf{w}^T\mathbf{g}(\mathbf{x}))^2 \mathbf{y} - \omega\mathbf{g}(\mathbf{x}) \right] \quad (\text{A.12})$$

and

$$F(\mathbf{x}, d\mathbf{x} / d\tau) = \mathbf{y}^T (d\mathbf{x} / d\tau) = \lambda \left[(2\mathbf{w}^T\mathbf{g}(\mathbf{x}))^2 \mathbf{y}^T\mathbf{y} - \omega\mathbf{y}^T\mathbf{g}(\mathbf{x}) \right] = \lambda (\omega^2 - \omega(\omega - 1)) = \lambda\omega \quad (\text{A.13})$$

where we have used Equation (A.11) and that $\omega = (1 + \mathbf{y}^T\mathbf{g}(\mathbf{x}))$ from Equation (A.9). Since by assumption $dt / d\tau = F(\mathbf{x}, d\mathbf{x} / d\tau) > 0$, λ must be always have the same sign as ω . We obtain the extremals of the variational problem by adding to the Equation (A.12) the second canonical equation:

$$\frac{d\mathbf{y}}{d\tau} = -\lambda \nabla_{\mathbf{x}} H(\mathbf{x}, \mathbf{y}) = -\lambda H_{\mathbf{x}}(\mathbf{x}, \mathbf{y}) = -\lambda \left[2(2\mathbf{w}^T \mathbf{g}(\mathbf{x}))(\mathbf{y}^T \mathbf{y}) \mathbf{H}\mathbf{w} - \omega \mathbf{H}\mathbf{y} \right] \quad (\text{A.14})$$

where \mathbf{H} is the Hessian of $V(\mathbf{x})$. For a solution, we have to use initial values $(\mathbf{x}_0, \mathbf{y}_0)$ for which the Hamiltonian $H(\mathbf{x}, \mathbf{y}) = 0$ holds in Equation (A.11). For many purposes it is more advantageous to introduce t as a parameter, whereby we set $dt / d\tau = F(\mathbf{x}, d\mathbf{x} / d\tau) = \lambda\omega = 1$. Moreover, if we replace the \mathbf{y} vector by its values from Equation (A.10), then the canonical equations become specially simplified. Namely, instead of Equation (A.12), we obtain

$$\frac{d\mathbf{x}}{dt} = -[\mathbf{I} - 2\mathbf{w}\mathbf{w}^T] \mathbf{g}(\mathbf{x}) \quad (\text{A.15})$$

and instead of Equation (A.14)

$$\frac{d\omega}{dt} \mathbf{w} + \omega \frac{d\mathbf{w}}{dt} = \omega \frac{\frac{d}{dt}(\mathbf{w}^T \mathbf{g}(\mathbf{x}))}{\mathbf{w}^T \mathbf{g}(\mathbf{x})} \mathbf{w} - \omega \mathbf{H}\mathbf{w}. \quad (\text{A.16})$$

We obtain from this equation by multiplying from the left by $[\mathbf{I} - \mathbf{w}\mathbf{w}^T]$ and using the idempotent property of this projector

$$\frac{d\mathbf{w}}{dt} = -[\mathbf{I} - \mathbf{w}\mathbf{w}^T] \mathbf{H}\mathbf{w}. \quad (\text{A.17})$$

This last equation forms with Equation (A.15) a system permitting all extremals of the GAD variational problem to be found. If we have calculated \mathbf{x} and \mathbf{w} vectors as functions of t , then we can determine the \mathbf{y} vector from the Equation (A.10) and that $\omega = (1 + \mathbf{y}^T \mathbf{g}(\mathbf{x})) = 2$.

Nevertheless, it is useful to note that ω can also be obtained by a quadrature, that is, from the equation

$$\frac{1}{\omega} \frac{d\omega}{dt} = 2 \left(\mathbf{w}^T \mathbf{H}\mathbf{w} - \frac{\mathbf{w}^T \mathbf{H}\mathbf{g}(\mathbf{x})}{\mathbf{w}^T \mathbf{g}(\mathbf{x})} \right) \quad (\text{A.18})$$

which arises by multiplying from the left by $\mathbf{w}\mathbf{w}^T$ Equation (A.16) and using Equations (A.15) and (A.17) and that $\mathbf{w}^T(d\mathbf{w} / dt) = 0$ due to normalization of the \mathbf{w} vector. Moreover, by differentiation of $\omega = (1 + \mathbf{y}^T \mathbf{g}(\mathbf{x}))$ and employing Equations (A.9), (A.13) and (A.14) we verify that

$$\frac{1}{\omega} \frac{d\omega}{dt} = \mathbf{w}^T \mathbf{H}\mathbf{w} - \frac{\mathbf{w}^T \mathbf{H}\mathbf{g}(\mathbf{x})}{\mathbf{w}^T \mathbf{g}(\mathbf{x})}. \quad (\text{A.19})$$

Equating the Equations (A.18) and (A.19) we obtain that $\omega^{-1} (d\omega / dt) = 0$ implying that $\omega = \text{constant}$, being this constant equal two along the GAD extremals. The right hand

side of Equations (A.18) and (A.19) is equal zero and coincides with the expression $d(\mathbf{y}^T \mathbf{g}(\mathbf{x})) / dt$ which is also zero since $\mathbf{y}^T \mathbf{g}(\mathbf{x}) = 1$ through the GAD extremal.

In order to decide whether a given line element is positive or negative regular, it is best to calculate the E -function itself

$$\begin{aligned} E &= F(\mathbf{x}, d\mathbf{x}'/dt) - (\nabla_{d\mathbf{x}/dt} F(\mathbf{x}, d\mathbf{x}'/dt))^T (d\mathbf{x}'/dt) \\ &= \lambda' \left(\omega' - \mathbf{y}^T \left((2\mathbf{w}'^T \mathbf{g}(\mathbf{x}))^2 \mathbf{y}' - \omega' \mathbf{g}(\mathbf{x}) \right) \right). \end{aligned} \quad (\text{A.20})$$

If in this, according to $\omega = (1 + \mathbf{y}^T \mathbf{g}(\mathbf{x}))$ and Equation (A.10), we set $\mathbf{y}^T \mathbf{g}(\mathbf{x}) = \omega - 1$, and

$$(2\mathbf{w}'^T \mathbf{g}(\mathbf{x}))^2 \mathbf{y}^T \mathbf{y}' = \omega' \omega \frac{\mathbf{w}'^T \mathbf{g}(\mathbf{x})}{\mathbf{w}'^T \mathbf{g}(\mathbf{x})} \mathbf{w}'^T \mathbf{w}' \quad (\text{A.21})$$

then we obtain the formula

$$E = \lambda' \omega' \omega \left(1 - \frac{\mathbf{w}'^T \mathbf{g}(\mathbf{x})}{\mathbf{w}'^T \mathbf{g}(\mathbf{x})} \mathbf{w}'^T \mathbf{w}' \right) = \lambda' \omega' \omega \mathbf{w}'^T \mathbf{E} \mathbf{w}' \quad (\text{A.22})$$

where

$$\mathbf{E} = \left[\mathbf{I} - \frac{1}{(2\mathbf{w}'^T \mathbf{g}(\mathbf{x}))} (\mathbf{g}(\mathbf{x}) \mathbf{w}'^T + \mathbf{w}' \mathbf{g}^T(\mathbf{x})) \right] \quad (\text{A.23})$$

since $\mathbf{w}'^T \mathbf{w}' = 1$. The sign of $\det(\mathbf{E})$ determines the sign of the E -function. The line elements are therefore all strong, they are positive or negative regular according the value of $\det(\mathbf{E})$ since $\lambda' \omega' > 0$ and $\omega = 2$. The only line elements which form an exception are the anomalous line elements for which $\mathbf{w}'^T \mathbf{g}(\mathbf{x}) = 0$. These line elements lie on the limit curves for the stationary field of extremals. The \mathbf{E} matrix of Equation (A.23) can be rewritten as

$$\mathbf{E} = \left\{ \mathbf{I} - [\mathbf{w}' | \mathbf{g}_n] \begin{bmatrix} 0 & 1/(2 \cos \beta) \\ 1/(2 \cos \beta) & 0 \end{bmatrix} [\mathbf{w}' | \mathbf{g}_n]^T \right\} \quad (\text{A.24})$$

where \mathbf{g}_n is the normalized gradient vector, $\mathbf{g}(\mathbf{x})$, and $\cos \beta = \mathbf{w}'^T \mathbf{g}_n$. The spectral decomposition of this \mathbf{E} matrix is

$$\mathbf{E} = \frac{\cos \beta - 1}{2 \cos \beta} \mathbf{v}_1 \mathbf{v}_1^T + \frac{\cos \beta + 1}{2 \cos \beta} \mathbf{v}_2 \mathbf{v}_2^T + \sum_{i=3}^N \mathbf{v}_i \mathbf{v}_i^T, \quad \mathbf{v}_i^T \mathbf{v}_j = \delta_{ij} \quad \text{for } i, j = 1, \dots, N \quad (\text{A.25})$$

where δ_{ij} is Kronecker's delta ($\delta_{ij} = 1$ for $i = j$ and 0 otherwise) and

$$\mathbf{v}_1 = \frac{1}{2} \sqrt{\frac{2}{1+\cos\beta}} (\mathbf{g}_n + \mathbf{w}); \quad \mathbf{v}_2 = \frac{1}{2} \sqrt{\frac{2}{1-\cos\beta}} (\mathbf{g}_n - \mathbf{w}). \quad (\text{A.26})$$

From Equation (A.25) we obtain that $\det(\mathbf{E}) = -1/4 \operatorname{tg}^2 \beta \leq 0$. The final expression of the E -function is

$$E = \lambda' \omega' \omega \left[1 + \left(\frac{\cos\beta - 1}{2\cos\beta} - 1 \right) (\mathbf{v}_1^T \mathbf{w}')^2 + \left(\frac{\cos\beta + 1}{2\cos\beta} - 1 \right) (\mathbf{v}_2^T \mathbf{w}')^2 \right] \quad (\text{A.27})$$

where we have used Equations (A.22) and (A.25) and that

$$\sum_{i=3}^N (\mathbf{v}_i^T \mathbf{w}')^2 = 1 - (\mathbf{v}_1^T \mathbf{w}')^2 - (\mathbf{v}_2^T \mathbf{w}')^2. \quad (\text{A.28})$$

In the case that $\mathbf{g}_n = \mathbf{w}$, then $\cos\beta = 1$, $\mathbf{v}_1 = \mathbf{g}_n$ and $\mathbf{v}_2 = \mathbf{0}$. In these conditions Equation (A.27) reduces to

$$E = \lambda' \omega' \omega [1 - \cos^2 \beta'] \quad (\text{A.29})$$

where $\cos\beta' = \mathbf{w}'^T \mathbf{g}_n$. Notice that now $E \geq 0$. The reason why $\det(\mathbf{E}) \leq 0$ except when $\mathbf{w} = \mathbf{g}_n$ can be found analyzing the structure of the indicatrix given in Equation (A.7), rearranged in the following way

$$(\mathbf{z} + \mathbf{g}(\mathbf{x}))^T (\mathbf{z} + \mathbf{g}(\mathbf{x})) = (2\mathbf{w}^T \mathbf{g}(\mathbf{x}))^2 \quad (\text{A.30})$$

where the left-hand side part depends of \mathbf{z} and \mathbf{x} while the right-hand side part depends of \mathbf{x} and the \mathbf{w} -vector. The indicatrix is the “wave front” of all extremal curves emerging from \mathbf{x} . If in this fixed point we vary the vector \mathbf{w} , then we obtain a new direction of progress but also the length of the “radius” vector of the indicatrix, $(\mathbf{z} + \mathbf{g}(\mathbf{x}))$, has varied where \mathbf{z} is a vector centered in the point $\mathbf{0}$ and ending in the indicatrix and related with the tangent vector, $d\mathbf{x} / dt$. We recall that $d\mathbf{x} / dt = \mathbf{z}$ since $F(\mathbf{x}, d\mathbf{x} / dt) = dt / dt = 1$ through the extremal curve. The square of the length of this “radius” vector is $4 \mathbf{g}^T \mathbf{g} \cos^2 \beta$ and its values are in between zero and $4 \mathbf{g}^T \mathbf{g}$, both included. However we can found at the same point \mathbf{x} a new vector \mathbf{w}' such that $\cos^2 \beta < \cos^2 \beta'$. For this new vector \mathbf{w}' the “radius” vector of the indicatrix is much greater that with the vector \mathbf{w} . This fact implies that the evolution of the point located at \mathbf{x} will go much further in the same Δt due to the relation between \mathbf{z} and the tangent vector. For this \mathbf{w}' we have $\mathbf{w}'^T \mathbf{E} \mathbf{w}' < 0$. The maximum “radius” vector of the indicatrix, $4 \mathbf{g}^T \mathbf{g}$, is attained when $\mathbf{w} = \mathbf{g}_n$ and for this reason $\mathbf{w}'^T \mathbf{E} \mathbf{w}' \geq 0$ for any \mathbf{w}' vector. In this case the curve evolves as steepest-ascent. Finally, when $\cos^2 \beta = 0$ the “radius” vector is equal to the zero vector and does not depend on the \mathbf{w} -vector. This is the anomalous element of the curve evolving as steepest-descent. As noted above, this anomalous element do not minimize or maximize the functional, it is a limit curve for the stationary field of extremals.

Appendix B: Extension of the shrinking dimer approach to the general GAD Zermelo model for any index n .

The extension of the algorithm to the general GAD Zermelo model for any n is straightforward. In this case the tangent vector has the form

$$\frac{d\mathbf{x}}{dt} = -[\mathbf{I} - 2\mathbf{W}\mathbf{W}^T]\mathbf{g}(\mathbf{x}) \quad (\text{B.1})$$

where the matrix \mathbf{W} is

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}^{(1)} & \dots & \mathbf{w}^{(n)} \end{bmatrix} \quad (\text{B.2})$$

being $n \leq N$ the dimension of the subset of control vectors and N the dimension of the problem. In addition it is $\mathbf{W}^T\mathbf{W} = \mathbf{I}_n$, the unit matrix of dimension n . The control matrix \mathbf{W} evolves according to the equation

$$\frac{d}{dt}\mathbf{W} = -[\mathbf{I} - \mathbf{W}\mathbf{W}^T]\mathbf{H}\mathbf{W}. \quad (\text{B.3})$$

Equations (B.1) and (B.3) are the Equations (1) and (2) emphasizing that $n > 1$. It is the case when more than one directional vector is involved. To solve Equations (B.1) and (B.3) using the dimer method at the i th iteration, first we define the two sets of points

$$\begin{aligned} \mathbf{x}_i^{(\mu)} &= \mathbf{x}_{c_i} - (1 - p_i^{(\mu)})d_i^{(\mu)}\mathbf{w}_i^{(\mu)} \\ \mathbf{x}_{i+1}^{(\mu)} &= \mathbf{x}_{c_i} + p_i^{(\mu)}d_i^{(\mu)}\mathbf{w}_i^{(\mu)} \end{aligned} \quad \mu = 1, \dots, n \quad (\text{B.4})$$

Note that μ is an index, not an exponent. In this way we have n dimers centered in the point \mathbf{x}_{c_i} , and the orientation of each dimer is given by the normalized control vector $\mathbf{w}_i^{(\mu)}$, $\mathbf{x}_{i+1}^{(\mu)} - \mathbf{x}_i^{(\mu)} = d_i^{(\mu)}\mathbf{w}_i^{(\mu)}$. As in the simple case, the approximated gradient at the inner point of each dimer is, $\mathbf{g}(\mathbf{x}_{c_i}) = (1 - p_i^{(\mu)})\mathbf{g}(\mathbf{x}_i^{(\mu)}) + p_i^{(\mu)}\mathbf{g}(\mathbf{x}_{i+1}^{(\mu)})$. We define the set of vectors

$$\mathbf{k}_i^{(\mu)} = \left(\mathbf{g}(\mathbf{x}_{i+1}^{(\mu)}) - \mathbf{g}(\mathbf{x}_i^{(\mu)}) \right) / d_i^{(\mu)} \quad \mu = 1, \dots, n. \quad (\text{B.5})$$

With these definitions we solve Equation (B.3)

$$\mathbf{w}_{i+1}^{(\mu)} = \mathbf{w}_i^{(\mu)} - \Delta t_i [\mathbf{I} - \mathbf{W}_i\mathbf{W}_i^T] \mathbf{k}_i^{(\mu)} \quad \mu = 1, \dots, n \quad (\text{B.6})$$

where \mathbf{W}_i is the \mathbf{W} matrix at the i th iteration. For the n -dimer translation we solve Equation (B.1) using the approximation

$$\mathbf{x}_{c\ i+1} = \mathbf{x}_{c\ i} - \Delta t_i \left[\mathbf{I} - 2\mathbf{W}_i \mathbf{W}_i^T \right] \mathbf{g}(\mathbf{x}_{c\ i}) \quad (\text{B.7})$$

Proceeding as in the previous case, we collect the vectors, $\mathbf{x}_{c\ i}$ and $\{\mathbf{w}_i^{(\mu)}\}_{\mu=1}^n$, into the vector \mathbf{h}_i^{ext} , namely, $\mathbf{h}_i^{ext T} = (\mathbf{x}_{c\ i}^T, \mathbf{w}_i^{(1)T}, \dots, \mathbf{w}_i^{(n)T})$, and the residues into the vector \mathbf{r}_i^{ext}

$$\mathbf{r}_i^{ext} = \begin{pmatrix} \left[\mathbf{I} - 2\mathbf{W}_i \mathbf{W}_i^T \right] \mathbf{g}(\mathbf{x}_{c\ i}) \\ \left[\mathbf{I} - \mathbf{W}_i \mathbf{W}_i^T \right] \mathbf{k}_i^{(1)} \\ \vdots \\ \left[\mathbf{I} - \mathbf{W}_i \mathbf{W}_i^T \right] \mathbf{k}_i^{(n)} \end{pmatrix} \quad (\text{B.8})$$

Finally with these definitions, the set of Equations (B.1) and (B.3) can be written in a compact form

$$\Delta \mathbf{h}_i^{ext} = -\Delta t_i \mathbf{r}_i^{ext} \quad (\text{B.9})$$

where $\Delta \mathbf{h}_i^{ext} = \mathbf{h}_{i+1}^{ext} - \mathbf{h}_i^{ext}$. The application of the quasi-Newton property, $\Delta \mathbf{h}_{i-1}^{ext} = -\Delta t_i \Delta \mathbf{r}_{i-1}^{ext}$ where $\Delta \mathbf{r}_{i-1} = \mathbf{r}_i^{ext} - \mathbf{r}_{i-1}^{ext}$, follows on a parallel way. We note that in the present case the normalization of the set of vectors $\{\mathbf{w}_i^{(\mu)}\}_{\mu=1}^n$ at each iteration can be done by either the Gram-Schmidt [29] or Löwdin [30] procedures. We are using the Löwdin procedure.

A refinement of the above algorithm consists in to replace the term $(\mathbf{g}(\mathbf{x}_{i+1}) - \mathbf{g}(\mathbf{x}_i)) / d_i$ appearing in Equation (5) or the set of vectors, $\{\mathbf{k}_i^{(\mu)}\}_{\mu=1}^n$, in Equation (B.6) by $\mathbf{H}_i \mathbf{w}_i$ or by the set of vectors, $\{\mathbf{H}_i \mathbf{w}_i^{(\mu)}\}_{\mu=1}^n$, respectively. The \mathbf{H}_i represents the Hessian matrix updated at the i th iteration. The GAD Zermelo path walks through the PES visiting regions where the Hessian matrix may or may not be positive definite. For this reason it is used the update formula of Murtagh-Sargent-Powell [31, 32]

$$\mathbf{H}_i = \mathbf{H}_{i-1} + \frac{\mathbf{j}_{i-1} \Delta \mathbf{x}_{c\ i-1}^T \mathbf{M}_{i-1} + \mathbf{M}_{i-1} \Delta \mathbf{x}_{c\ i-1} \mathbf{j}_{i-1}^T}{\Delta \mathbf{x}_{c\ i-1}^T \mathbf{M}_{i-1} \Delta \mathbf{x}_{c\ i-1}} - \left(\mathbf{j}_{i-1}^T \Delta \mathbf{x}_{c\ i-1} \right) \frac{\mathbf{M}_{i-1} \Delta \mathbf{x}_{c\ i-1} \Delta \mathbf{x}_{c\ i-1}^T \mathbf{M}_{i-1}}{\left(\Delta \mathbf{x}_{c\ i-1}^T \mathbf{M}_{i-1} \Delta \mathbf{x}_{c\ i-1} \right)^2} \quad (\text{B.10})$$

where, $\mathbf{j}_{i-1} = \mathbf{g}(\mathbf{x}_{c\ i}) - \mathbf{g}(\mathbf{x}_{c\ i-1}) - \mathbf{H}_{i-1} \Delta \mathbf{x}_{c\ i-1} = \mathbf{g}(\mathbf{x}_{c\ i}) - \mathbf{g}(\mathbf{x}_{c\ i-1}) - \mathbf{H}_{i-1} (\mathbf{x}_{c\ i} - \mathbf{x}_{c\ i-1})$ and $\mathbf{M}_{i-1} = a_{i-1} \mathbf{I} + (1 - a_{i-1}) (\Delta \mathbf{x}_{c\ i-1}^T \Delta \mathbf{x}_{c\ i-1}) / (\mathbf{j}_{i-1}^T \Delta \mathbf{x}_{c\ i-1})^2 \mathbf{j}_{i-1} \mathbf{j}_{i-1}^T$ being $1 \geq a_{i-1} \geq 0$. A version of this update Hessian formula was also reported for large dimensional systems [33]. It is well known that update formulae like Equation (B.10) are numerical stable even when the dimension of the system is large [29, 34].

Appendix C: Proof of the validity of Equation 4.

A PES, $V(\mathbf{x})$, is considered a continuous function in \mathbf{x} , at least three times differentiable with respect to \mathbf{x} and its derivatives should be continuous. In other words, a smooth function of many variables can be approximated by a Taylor series around any point. The basic idea is that in the smooth functions it is true that the Hessian matrix, $\mathbf{H}(\mathbf{x})$, maps differences in position into differences in gradient, $\mathbf{g}(\mathbf{x}') - \mathbf{g}(\mathbf{x}) = \mathbf{H}(\mathbf{x}) (\mathbf{x}' - \mathbf{x}) + O(\|\mathbf{x}' - \mathbf{x}\|^2)$, where $\|\cdot\|$ means norm of a vector in Euclidean metric and $O(\|\mathbf{x}' - \mathbf{x}\|^2)$ means that as \mathbf{x}' approximates to \mathbf{x} then $\|\mathbf{g}(\mathbf{x}') - \mathbf{g}(\mathbf{x}) - \mathbf{H}(\mathbf{x}) (\mathbf{x}' - \mathbf{x})\| \leq c \|\mathbf{x}' - \mathbf{x}\|^2$ where c is a positive constant. The $O(\cdot)$ notation describes the limiting behavior of a function when the variable tends toward a particular value. Now we take $\mathbf{x}' = \mathbf{x}_{i+1}$, $\mathbf{x} = \mathbf{x}_i$ and we represent the differences between these two points as $\mathbf{x}_{i+1} - \mathbf{x}_i = \mathbf{w}_i d_i$, where \mathbf{w}_i is a normalized vector. In this way we can write, $\mathbf{g}(\mathbf{x}_{i+1}) - \mathbf{g}(\mathbf{x}_i) = \mathbf{H}(\mathbf{x}_i) \mathbf{w}_i d_i + O(d_i^2)$, consequently any point in between \mathbf{x}_{i+1} and \mathbf{x}_i satisfies this equality in particular $\mathbf{x}_{ci} = (1 - p_i) \mathbf{x}_i + p_i \mathbf{x}_{i+1}$ where $0 \leq p_i \leq 1$. With these notations and definitions at hand we proof Equation (4). First, we rearrange Equation (4) as follows,

$$\mathbf{g}(\mathbf{x}_{ci}) - \mathbf{g}(\mathbf{x}_i) = p_i(\mathbf{g}(\mathbf{x}_{i+1}) - \mathbf{g}(\mathbf{x}_i)) = p_i (\mathbf{H}(\mathbf{x}_i) \mathbf{w}_i d_i + O(d_i^2)) = p_i \mathbf{H}(\mathbf{x}_i) \mathbf{w}_i d_i + O(d_i^2). \quad (\text{C.1})$$

Second, from the definition of the point \mathbf{x}_{ci} , we can write, $\mathbf{x}_{ci} - \mathbf{x}_i = p_i (\mathbf{x}_{i+1} - \mathbf{x}_i) = p_i \mathbf{w}_i d_i$. Now,

$$\mathbf{g}(\mathbf{x}_{ci}) - \mathbf{g}(\mathbf{x}_i) = \mathbf{H}(\mathbf{x}_i) (\mathbf{x}_{ci} - \mathbf{x}_i) + O(\|\mathbf{x}_{ci} - \mathbf{x}_i\|^2) = p_i \mathbf{H}(\mathbf{x}_i) \mathbf{w}_i d_i + O(p_i^2 d_i^2). \quad (\text{C.2})$$

We note that $\|\mathbf{x}_{ci} - \mathbf{x}_i\|^2 = p_i^2 d_i^2 \leq \|\mathbf{x}_{i+1} - \mathbf{x}_i\|^2 = d_i^2$, since $0 \leq p_i \leq 1$ with equality if $p_i = 1$. Defined in this way, the \mathbf{x}_{ci} is an interior point of the ball centered in \mathbf{x}_i with radius d_i . Since $p_i^2 d_i^2 \leq d_i^2$ then the limiting behavior $O(d_i^2)$ is less restricted than $O(p_i^2 d_i^2)$. Comparing Equations (C.1) and (C.2) we see that they are identical differing only in the limiting behavior $O(d_i^2)$. This concludes the proofs of the general validity of Equation (4) for any point of the PES within the quadratic order of the expansion with respect to a selected point.

References

1. P. G. Mezey, Potential Energy Hypersurfaces, Elsevier, NewYork, 1987
2. D. Heidrich, The Reaction Path in Chemistry: Current Approaches and Perspectives, Kluwer, Dordrecht, 1995
3. I. Kolossváry, W. C. Guida, C. J. Am. Chem. Soc., **118**, 5011 (1996)
4. G. Crippen, H. A. Scheraga, Arch. Biochem. Biophys., **144**, 462 (1971)
5. S. Nakamura, H. Hirose, M. Ikeguchi, J. Doi, J. Phys. Chem., **99**, 8374 (1995)
6. G. T. Barkema, N. Mousseau, Phys. Rev. Lett., **77**, 4358 (1996)
7. N. Mousseau, G. T. Barkema, Phys. Rev. B, **61**, 1898 (2000)
8. E. Cancès, F. Legoll, M.-C. Marinica, K. Minoukadeh, F. Willaime, J. Chem. Phys., **130**, 114711 (2009)
9. W. E, X. Zhou, Nonlinearity, **24**, 1831 (2011)
10. A. Samanta, W. E, J. Chem. Phys., **136**, 124104 (2012)
11. J. M. Bofill, W. Quapp, M. Caballero, Chem. Phys. Lett., **583**, 203 (2013)
12. W. Quapp, J. M. Bofill, Theor. Chem. Acc., **133**, 1510 (2014)
13. W. Quapp, J. M. Bofill, E. Bernuz, J. Math. Chem., **53**, 41 (2015)

14. J. M. Bofill, W. Quapp, *Theor. Chem. Acc.*, **135**, 11 (2016)
15. G. Henkelman, H. Jonsson, *J. Chem. Phys.*, **111**, 7010 (1999)
16. E. Zermelo, *Z. Angew. Math. Mech.*, **11**, 114 (1931)
17. C. Carathéodory, *Variationsrechnung und partielle Differentialgleichungen erster Ordnung.*, B. G. Teubner, Berlin, 1935
18. J. M. Bofill, W. Quapp, *An Application of the Maximum Principle in Chemistry: A Method to Locate Transition States*, in *Theoretical & Quantum Chemistry at the Dawn's End of 21st Century*, edited by R. Carbó-Dorca & T. Chakraborty, Taylor & Francis Group, London, 2016
19. J. Barzilai, J. M. Borwein, *IMA J. Numer. Anal.*, **8** 141 (1988)
20. L. Zhang, Q. Du, Z. Zheng, *SIAM J. Sci. Comput.*, **38**, A528 (2016)
21. J. Zhang, Q. Du, *SIAM J. Numer. Anal.*, **50**, 1899 (2012)
22. S. Wolfe, H. B. Schlegel, I. G. Csizmadia, F. Bernardi, *J. Am. Chem. Soc.*, **97**, 2020 (1975)
23. W. Quapp, *J. Chem. Phys.*, **122**, 174106 (2005)
24. R. Olender, R. Elber, *J. Chem. Phys.*, **105**, 9299 (1996)
25. P. G. Bolhuis, D. Chandler, C. Dellago, P. L. Geissler, *Ann. Rev. Phys. Chem.*, **59**, 291 (2002)
26. C. Dellago, P. G. Bolhuis, P. L. Geissler, *Adv. Chem. Phys.*, **123**, 1 (2002)
27. A. Laio, M. Parinello, *Proc. Nat. Acad. Sci.*, **99**, 12562 (2002)
28. C. Junghans, D. Perez, T. Vogel, *J. Chem. Theory Comput.*, **10**, 1843 (2014)
29. M. Hestenes, *Conjugate Direction Methods in Optimization*, Springer-Verlag, New York, Heidelberg, Berlin, 1980
30. P.-O. Löwdin, *Adv. Phys.*, **5**, 1 (1956)
31. J. M. Bofill, *J. Comput. Chem.*, **15**, 1 (1994)
32. J. M. Bofill, M. Comajuan, *J. Comput. Chem.*, **16**, 1326 (1995)
33. J. M. Anglada, E. Besalú, J. M. Bofill, J. Rubio, *J. Math. Chem.*, **25**, 85 (1999)
34. R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, New York, 1987
35. Z. Smedarchina, W. Siebrand, A. Fernández-Ramos, *J. Chem. Phys.*, **127**, 174513 (2007)
36. G. Albareda, J. M. Bofill, I. Tavernelli, F. Huarte-Larrañaga, F. Illas, A. Rubio, *J. Phys. Chem. Lett.*, **6**, 1529 (2015)
37. H. Mühlenbein, D. Schomisch, J. Born, *Parallel. Comp.*, **17**, 619 (1991)
38. M. F. Gelin, D. S. Kosov, *J. Chem. Phys.*, **126**, 244501 (2007)