A growing string method for the reaction pathway defined by a Newton trajectory

WOLFGANG QUAPP

UNIVERSITÄT LEIPZIG

Mathematical Institute

quapp@rz.uni-leipzig.de
www.math.uni-leipzig.de/∼quapp/GS

The reaction path is an important concept of theoretical chemistry. We use a projection operator for the following of the Newton trajectory (NT) along the reaction valley of the potential energy surface. We describe the numerical scheme for the string method adapting the growing string (GS). The combination of the Newton projector and the growing string idea is an improvement of both methods, and a great saving of the number of iterations needed to find the pathway over the saddle point. This combination GS-NT is at the best of our knowledge new. We employ two different corrector methods: first the use of projected gradient steps, and second a conjugated gradient method, the CG+ method of Liu, Nocedal and Waltz, generalized by projectors.

Examples are Lennard-Jones clusters LJ_{22}, and an N-methyl-alanyl-acetamide (alanine dipeptide) rearrangement between the minima $C7_{ax}$ and C5. For the latter, the growing string calculation is interfaced with the Gassian03 package.

Ref.: The Journal of Chemical Physics 122 Issue 17 (2005) 174106.

I. INTRODUCTION

• The design of a robust method for the determination of a reaction pathway (RP) on a complex energy landscape is a very important problem. We use Newton trajectories (NT) in double-ended methods[1] combined with a proposal of a growing string (GS) method recently given by B. Peters et al.[2]

• The RP of an adiabatic potential energy surface (PES) is the usual approach to the theoretical kinetics of larger chemical systems. It is any one-dimensional line connecting two minima by passing the saddle point (SP) in n-dimensional coordinate space.

• We do not find it difficult to recognise RPs, we find it difficult indeed to offer a definition that is conceptually watertight, and immune to counter-example.[3]

• We here use the reduced gradient following (RGF),[4,5] also called NT.
We insist that the search of an appropriate RP is not necessarily equivalent to the finding of the steepest descent from SP.

• Most RPs can be defined with the help of projection operators. The tool is employed in string methods: the string is divided into a collection of nodes which are moved by projectors. The nodes $\mathrm{x}_1, \ldots, \mathrm{x}_m$ represent the RP by a chain of length m where the endpoints may be the minima x_0 and x_{m+1}.

• Our GS-NT method is divided into predictor and corrector steps.
We describe both, and especially the termination criterion and performance of the corrector, and at the very end, the implementation of the corrector step.

II. NEWTON TRAJECTORIES

• Projection operator: Choose an n-dimensional column vector $\mathbf{r}$ for the projection. It has to be a unit vector. The transposed vector $\mathbf{r}^T$ is a row vector. The dyadic product $\mathbf{D_r} = \mathbf{r} \cdot \mathbf{r}^T$ is an $(n \times n)$ matrix. $\mathbf{D_r}$ projects with $\mathbf{r}$:

$$\mathbf{D_r}\ \mathbf{r} = (\mathbf{r} \cdot \mathbf{r}^T) \cdot \mathbf{r} = \mathbf{r}\ (\mathbf{r}^T \cdot \mathbf{r}) = \mathbf{r}\ . \tag{1}$$

The projector which projects down the $\mathbf{r}$, is with the unit matrix $\mathbf{I}$

$$\mathbf{P_r} = \mathbf{I} - \mathbf{D_r}\ . \tag{2}$$

The concept of RGF[4,5] is that a selected gradient direction is fixed along the curve $\mathbf{x}(s)$ with curve parameter s for gradient $\mathbf{g}$ of the PES

$$\mathbf{g}(\mathbf{x}(s))/||\mathbf{g}(\mathbf{x}(s))|| = \mathbf{r}\ . \tag{3}$$

Which direction $\mathbf{r}$ is to select is a certain arbitrariness.[1]

• The original driven coordinate method employs the eigenvector direction of a reaction valley. However, in the (higher dimensional) examples below, we will not execute the calculation of the Hessian, at all.

• Property (3) is realizable by a projection of the gradient employing $\mathbf{P_r}$ of (2). We pose the Newton projector[5]

$$\mathbf{P_r}\,\mathbf{g}(\mathbf{x}(s)) = \mathbf{0}\ . \tag{4}$$

$\mathbf{P_r}$ is an $n \times n$ matrix of rank $n-1$. The solution $\mathbf{x}$(s) is named Newton trajectory. If starting at a minimum, eq.(4) is trivially fulfilled for every direction $\mathbf{r}$. Thus, we may choose any direction because there is a solution which starts at the minimum. A family of NTs connects the extrema if we vary the search direction $\mathbf{r}$.[6] Projector (4) can be used in a string method at every actual chain point, without any further derivative.[1] If it does not result in zero, choose the downhill direction of a node moving along

$$\mathbf{p} = -(\mathbf{I} - \mathbf{r} \cdot \mathbf{r}^T)\ \mathbf{g}\ . \tag{5}$$

The application of the projector does not need the tangent of the curve, it does not need any reparameterization of the string, it does not need any spring forces of the chain, and every chain point can be moved independently.[1]

It predestinates the Newton projector for the growing string method[2] in an exceptional kind. The main problem under an application of operator (5) is nothing but finding an appropriate steplength, or a dampening factor η for $\mathbf{p}$.

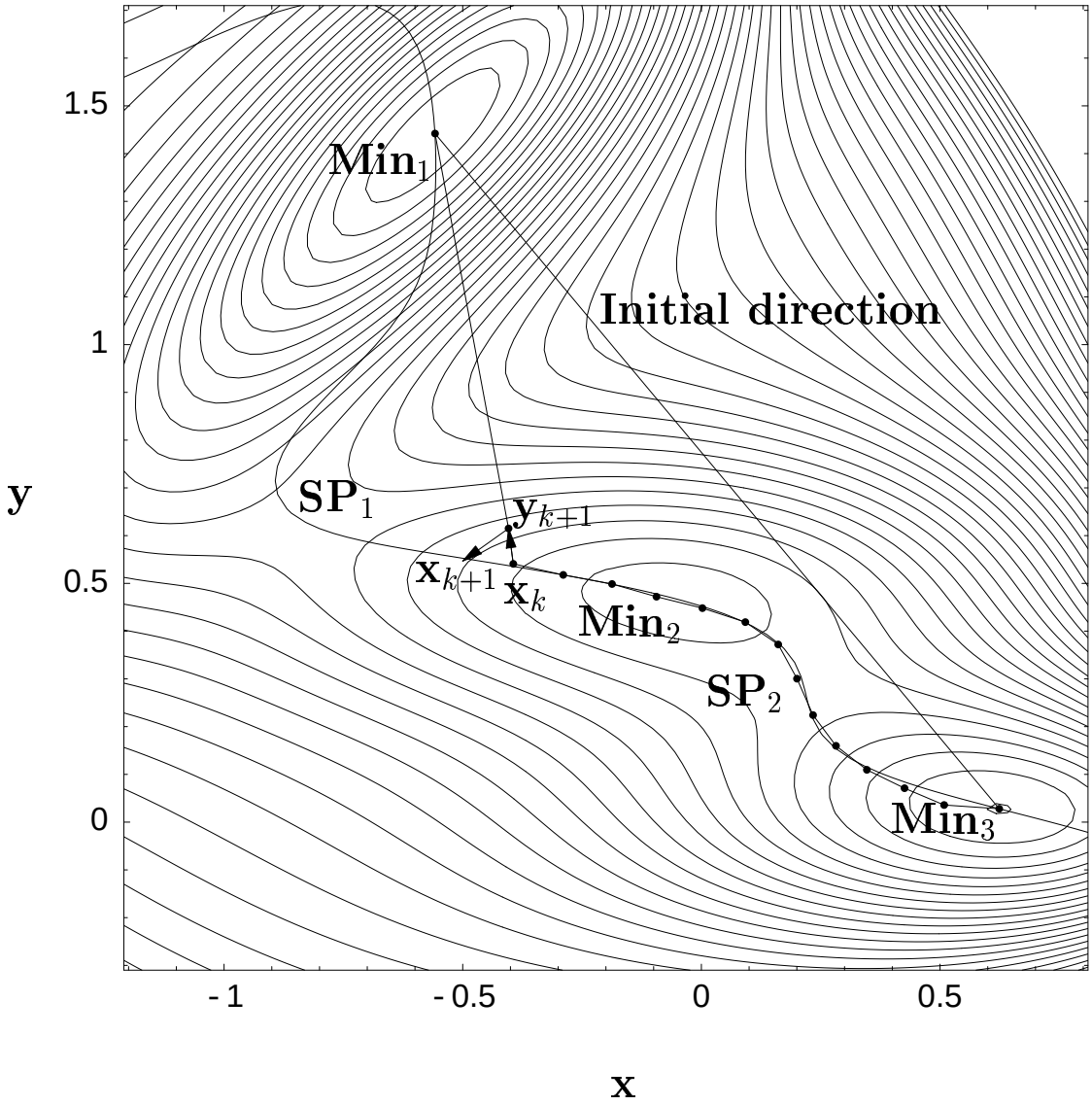


FIG. 1. GS method on Müller-Brown potential from minimum M_3 to minimum M_1. The predictor step at $\mathbf{x}_{13}$ is shown, and the corrector back to the searched RP (arrows). The desired number of nodes is L=23. The Newton trajectory to initial direction is included for comparison. The growing string follows the NT very well.

III. GROWING STRING METHOD

The growing string method adaptively evolves the string from its endpoint(s). We search an RP which should connect the initial minimum, $\mathbf{x}_{ini}$, with the end $\mathbf{x}_{fin}$ by a chain of m nodes $\mathbf{x}_k$. We calculate successive nodes beginning at the initial minimum.

(i) $\mathbf{x}_k$ is an approximated node on the RP. We choose a next guess point, $\mathbf{y}_{k+1}$, of the string between the actual node and the final minimum by

$$\mathbf{y}_{k+1} = \lambda\, \mathbf{x}_k + (1-\lambda)\ \mathbf{x}_{fin}\ , \quad \lambda = \frac{m-k}{m+1-k}\ , \quad k = 0, .., m-1, \tag{6}$$

see Fig.1, where m is the desired number of nodes, and $\mathbf{x}_0$=$\mathbf{x}_{ini}$, $\mathbf{x}_{m+1}$=$\mathbf{x}_{fin}$.

(ii) Doing corrector steps with the gradient at the guess point $\mathbf{y}_{k+1}$ using the Newton projector (5) up to convergence, thus up to a threshold ϵ for the right side of (4), see Fig.1: the steps go orthogonally to the search direction. We use two different corrector methods: first, the pure Newton projector with an appropriate dampening factor, η, and second, a conjugate gradient method modified by the Newton projector.

(iii) The point at convergence is the next node, $\mathbf{x}_{k+1}$, and the process is repeated up to the final minimum.

• To execute the instruction, we need the desirable (and fixed) number of nodes, m, and the search direction, $\mathbf{r}$. The former results in the steplength for the guess point, the latter in the direction of the corrector to the NT (modelling the RP).

• To check the convergence of $\mathbf{p}$, that the projected and inverse gradient $\mathbf{g}$ in eq.(5) is zero, we also need a convergence criterion, ϵ.

• We do not reparametrize the string. The NT with respect to $\mathbf{r}$ should also be followed if the nodes increase their distance and are not distributed evenly. Generally, we hope for an automatically evenly distributed growing string which then is an indicator for a well chosen search direction and a good execution of the corrector.

IV. IMPROVEMENT OF THE PROJECTED GRADIENT SEARCH BY THE CG+ METHOD

• The corrector step by eq. (5) uses a projected gradient, which works well in low dimensions, and which works finely if the PES section of the main direction is not too flat, in relation to further orthogonal directions. However, the steplength of gradient methods is always a problem. Using a dampening factor, η, one can, with some experience, successfully apply the steps $\mathbf{p}$ of eq. (5). One condition is that the tolerance for the corrector is not too sharp! But this is recommendable because we search a coarse approximation of the RP which should lead over a maximum value of the energy in any neighborhood of the SP. The SP itself can then be improved by another method.

• In the higher dimensional case, however, we need improved methods. A minimizer with high merits is the CG+ algorithm of Nocedal et al.[8] CG+ is a Conjugate Gradient code used for solving nonlinear, unconstrained optimization problems. The CG+ routine is especially effective on problems involving a large number of variables. It needs the subroutine which, given an input vector, $\mathbf{x}$, returns the function and gradient for the function one wants to minimize. The Hessian is never computed. The steplength along a search direction is additionally determined by a line search routine, which is a slight modification of a routine written by Moré and Thuente.[9] The purpose is to find a step which satisfies a sufficient decrease condition and a curvature condition.

• If applying the projected and inverse gradient $\mathbf{p}$ of eq. (5), in the steepest descent search, we go trivially orthogonally to the search direction $\mathbf{r}$ in every step. However, if applying CG+ like a minimization, using the projected and inverse gradient (5) for the input, this does not guarantee that the method moves in the constrained plane orthogonally to the search direction of the Newton trajectory search. We choose a more indirect way to safe the orthogonal search to the direction $\mathbf{r}$:

V. LAGRANGIAN CONDITION

We formally pose an additional degree of freedom to the energy function $E(\mathbf{x})$ by

$$L(\mathbf{x}, \lambda) = E(\mathbf{x}) - \lambda \, \mathbf{r}^T \cdot (\mathbf{x} - \mathbf{y}) \ , \tag{7}$$

and search a minimum for ($\mathbf{x}$, λ), where $\mathbf{y}$ may be the initial point of a corrector loop, see Fig.1. λ is the Lagrangian multiplier, and the linear equation of a hyperplane

$$C(\mathbf{x}) = 0 = \mathbf{r}^T \cdot (\mathbf{x} - \mathbf{y}) \tag{8}$$

should be fulfilled in the corrector loop, as well as the value of the (n+1st) variable

$$\lambda = \pm ||\mathbf{g}(\mathbf{x})|| \ . \tag{9}$$

The gradient for L is for the first n dimensions $\mathbf{g}(\mathbf{x}) - \lambda \, \mathbf{r}$, and the last entry is $-C(\mathbf{x})$. An application of $\mathbf{P_r}$ on the first n dimensions of the gradient of L again results in eq.(4), and eq.(3) is fulfilled with ansatz (9). We will find a minimum solution of the constrained problem (7) by employing the known conditions of L directly in a modified, but "unconstrained" CG+ run:

(i) We modify the procedure putting for every iteration step the "true" λ value (9) for the input.

(ii) We use the gradient of $L(\mathbf{x}, \lambda)$ for the CG+ program, however, the calculated steps $D_j, \ j = 1, ..$ of that procedure are projected by $\mathbf{P_r}$ in every loop. Thus, we only use steps in the hyperplane $C(\mathbf{x}) = 0$.

(iii) We suppress some strong tests of the CG+ program, for example, we allow that in some steps the search direction is not always a strong descent direction: we leap over the test $D_j^T \cdot \mathbf{g} > 0$. We generally restrict the steplength along a search direction D_j.

We found that this modified CG+ better searches in the projected subspace orthogonal to $\mathbf{r}$ than the pure projected gradient, $\mathbf{p}$. An example below needs a very sharp threshold of the corrector: here the modified CG+ works well.

VI. MÜLLER-BROWN PES

We use the Müller-Brown (MB) PES,[10] see Fig.1, for a test. We start at M_3 with a straight line between the two outer minima, M_3 and M_1. It is the initial search direction $\mathbf{r}$. We use the dampening factor of η=0.13, and the same convergence criterion of ϵ=0.08 for the coarse convergence of the loop of the actual guess point. We move every new guess point in its own loop. Note that the guess point is already near the RP: usually, the convergence of the corrector needs one or two steps. The dampening factor, η, can variate. However, going to higher values it is restricted by zigzagging of the algorithm.

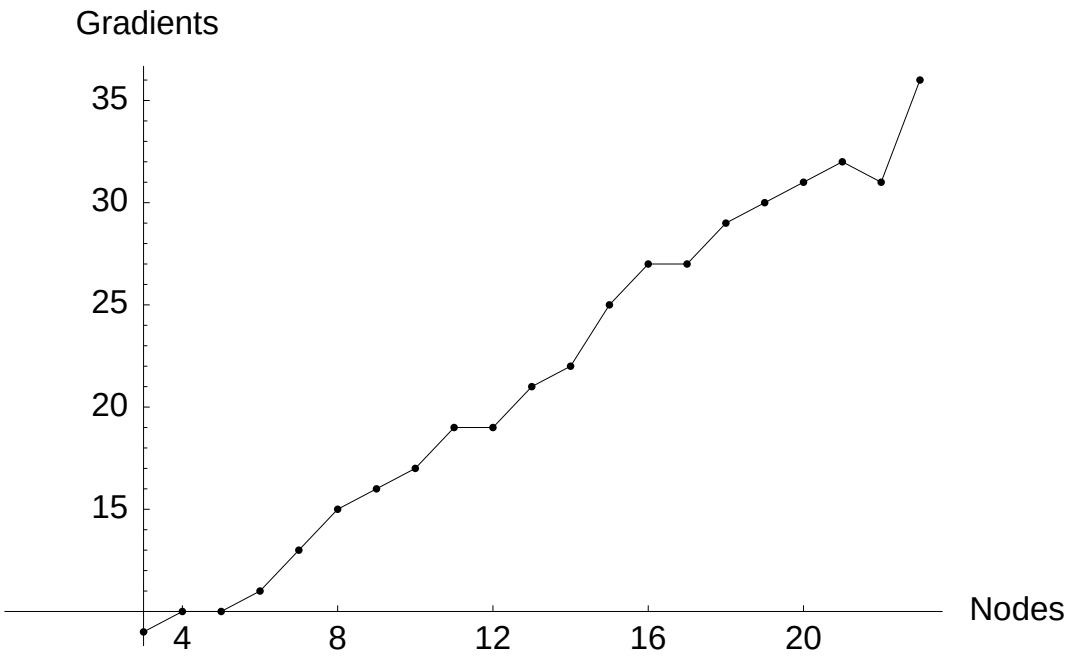


FIG. 2. Effort of the growing ends string method for the MB potential

The Figure shows the effort for different chain lengths. The effort increases quite linearly in the number of nodes from 9 gradient calculations which we need for 3 nodes to 36 gradients for 23 nodes. The improvement against the moving of a full initial chain[1] is dramatic, as well as the saving against the growing string method for the steepest descent pathway,[2] where the effort is between 40 and 100 gradient calculations.

VII. LENNARD-JONES CLUSTER LJ_{22}, CG+ METHOD

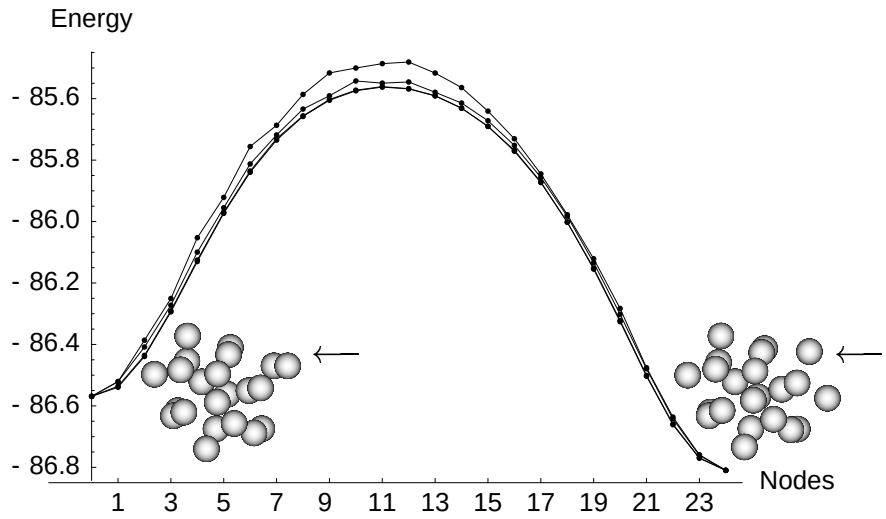


FIG. 3. Growing ends string method with 23 nodes for the LJ_{22} potential (assuming Argon). Shown is the energy profile. The left-arrows indicate the two atoms which are mainly involved into the rearrangement. From top to bottom the resulting curves for the corrector thresholds ϵ=1.0, 0.5, 0.1 are shown. The SP at node #11 is well approximated.

Test of CG+ method is given in Fig. 3. The global minimum of that cluster is from the Wales tables,[11] at the right hand side, and a neighbor second-lowest-energy minimum being only slightly higher in its energy. A chain of 23 nodes is used. The start chain is again the simple linear combination of the two minima in Cartesian coordinates. Successful runs are done with the different parameter,

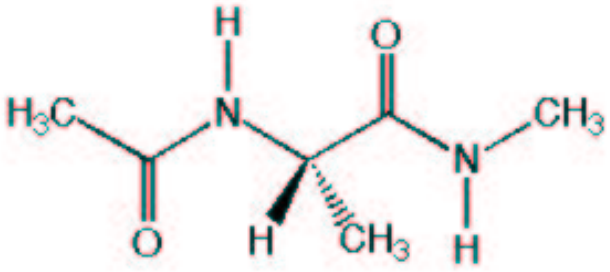
• the threshold ϵ=1.0, 0.5, 0.1, and 0.05, (now for every gradient component).

The CG+ method needs 56, 100, 212, or 328 gradient calculations, respectively. The last two runs are not to distinguish, they coincide in the graphics. The highest point of this lowest profile is exactly the SP of a rearrangement of the cluster. Also the first test with the coarse ϵ=1.0 results in a good estimate of the SP. The 56 gradients, which one needs in this most optimistic run, are less than the dimension of the problem. It means 2-3 gradients per node.

• The calculation of the LJ_{22} energies and gradients in such a number needs some seconds using a PC.

VIII. ALANINE DIPEPTIDE

For large peptide systems, we are still limited to the determination of one or a few RPs from a given minimum to another one. Which one is the global MEP: we do not know! If a search direction is chosen, and if the search is successful, we will obtain a "local" MEP. If the SP of the path is sufficiently low then the path may be the global MEP between the minima.



Scheme 1 Alanine dipeptide

• A common benchmark for testing protein modeling algorithms is the 66-dimensional PES of alanine dipeptide.

• We search a rearrangement between the minimum C5 and the $C7_{ax}$ minima. It is mainly a rotation along the (Φ , Ψ) angles from (-170,170)=(190,-190) to (75,-60) set of values of the backbone dihedral angles Φ (C-N-C_α-C) and Ψ (N-C_α-C-N).
C_α is the central C atom in Scheme 1.

• For a demonstration of the method, guess (6) is a provoking task, a challenge. Because, the growing string method avoids the very unrealistic structures of the "molecule" on the linear interpolation pathway. There the energies would be extremely high, some hundreds to thousand kcal/mol, where, on the possible rearrangement path, the SP is in the range of 9.5 kcal/mol above the lower minimum, C5.

• If we use the 3-21G level of computation, the example has a further drawback: the RP between the minima C5 and $C7_{ax}$ is a combined RP. There an intermediate minimum exists in between, and there two SPs exist between the minima C5 and $C7_{ax}$.

• A first series of computational tests are done with the corrector descent along the projected and inverse gradient (5). The predictor step may be given by (6); and the initial chain is between the C5 and the $C7_{ax}$ structure of alanine dipeptide.

• In contrast to the LJ_{22} example, the alanine dipeptide example is a different story. The projected gradient descent comes to its borderline here. Some results are illustrated in Fig. 4. The general steplength for corrector steps has to be dampened by a value of 0.6.

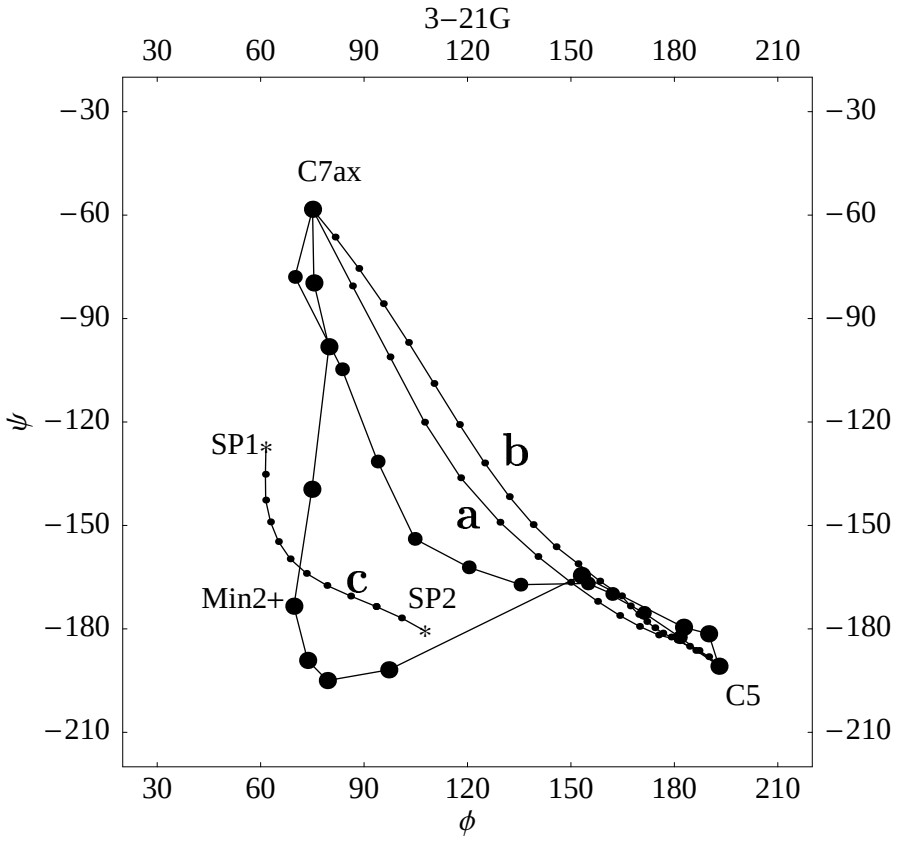


FIG. 4. Approximation of Newton trajectories for alanine dipeptide between the minima $C7_{ax}$ and C5. Methods: projected and inverse gradient (thin points) starting at C5, and modified CG+ optimization (thick bullets) starting at $C7_{ax}$. The two coordinates (Φ, Ψ) of the 60D internal coordinates are shown in a Ramachandran diagram. The "better" one CG+ result (more left and below) is to ϵ=0.001. The other chain belongs to ϵ=0.0075. The two SPs ($*$) and the intermediate minimum (+) are included.

(a) Optimistic gradient test: we chose ϵ=0.008 and 13 nodes between C5 and $C7_{ax}$. The norm of the reduced gradient will be accepted if it is smaller than ϵ, or if we have done a maximum of 55 corrector steps. In sum we need 604 gradient calculations for curve (a). The ϵ seems still a little too coarse, however, starting at the point with maximum energy of the chain obtained, one may execute the Berny optimization of the Gaussian03.[12]

- SP search in Gaussian03 by Berny's method.

```
$RunGauss
# scf=direct 3-21G opt (ts,saddle=1,CalcFC,noEigenTest) optcyc=99 test
```

which are followed by the geometry of the node. It leads to the SP after 70 steps. The SP_2 is situated at (Φ, Ψ)=(107.8^o, -180.7^o). Neighboring points of the chain of (a) lead to the intermediate minimum at (Φ, Ψ)=(64.3^o, -171.8^o) and to the SP_1 at (Φ, Ψ)=(61.5^o, -127.3^o), in the 3-21G basis.

(b) Pessimistic gradient test: in a contrary ansatz to (a), we use ϵ=0.000 5 and 23 nodes between C5 and $C7_{ax}$. The maximally allowed corrector steps are 151. The ϵ is so small that we cannot remain under it, and we cut the corrector in every step at the point after 151 steps. In total we need 23×151 gradient calculations for curve (b).
The Berny optimization process needs an eigenvalue of the Hessian of a negative signature. It is fulfilled here. Looking at Fig.4, it offers the speculation that node # 21, the node with maximal energy, is far away from SP_2. So, indeed, starting at the nearer node # 17 one already obtains the SP_2 in 63 Berny steps.

(c) The SPs obtained from curves (a) or (b) are used to start a further NT search for a connection curve in between. A result is shown in case (c). The curve is searched with the threshold ϵ=0.000 5, and a maximum of 151 steps (which is often exhausted). The curve indicates the intermediate minimum, but does not reach it.

The way out of the "pure" gradient-dilemma is a better descent routine:

CG+ method

Figure 4 shows results of CG+ calculations. The ϵ is used now to prove the smallness of every component of the gradient of the Lagrangian ansatz (7). The steplength for corrector steps is restricted by a value of 2.5 units. We additionally have changed the order of the two minima: we start with minimum $C7_{ax}$ and search a growing string to the C5 minimum. The reason is that we thus meet the lower SP_1 first. We use 10 nodes.

(i) Pessimistic CG+ calculation: The curve is calculated with threshold ϵ=0.001 for every component of the gradient, and a maximum of $2(n+1)$ steps per corrector. The maximal number of steps is throughout exhausted. (The energy shows a double hump shape. The profile is a very good approximation of an RP between the three included minima, and the two SPs in between.) In sum, it needs 1340 gradient calculations.

If we start an optimization by the Berny process of Gaussian03[12] at the first maximal node, #3, it needs 34 steps to converge to the flat SP_1, and using the second maximal node, #7, it needs 7 steps to converge immadiately to SP_2.

(ii) Optimistic CG+ calculation: The curve is calculated with threshold ϵ=0.007 5 for every component of the gradient. Every corrector loop has converged. The energy profile is not well adapted to the complicated situation of two SPs and an intermediate. But the profile is still a usable approximation of an RP between the start and final points. A SP is indicated by a maximum value of energy. In sum, the run needs 816 gradient calculations.

Fig. 5: since we know the existence of the intermediate minimum, we may use it in a last control calculation. We approximate in one run the NT to the direction between the C5 and the intermediate minimum, and start there, in a second run, to approximate the NT to the direction between the intermediate minimum and minimum $C7_{ax}$. Thus, we turn again the direction of the GS development, and start with C5.

10 nodes are fixed for every string and the convergence condition of the corrector is ϵ=0.005 for the stronger slopes on the first pathway, and ϵ=0.003 33 for the second pathway. Both values are between the cases used in Fig.4.

CG+ needs 540+420 gradient calculations, correspondingly.

The two GS-NT calculations need 9.6 hours using Gaussian03 for energy and gradient at an Itanium ®2 processor based HP workstation.

Most corrector loops converge on both pathways with less than 50 steps.

• We are compelled to emphasize that the approximations obtained seem to be perfect RPs. The mild edge at node 3 on the first pathway may be a shoulder of the PES.

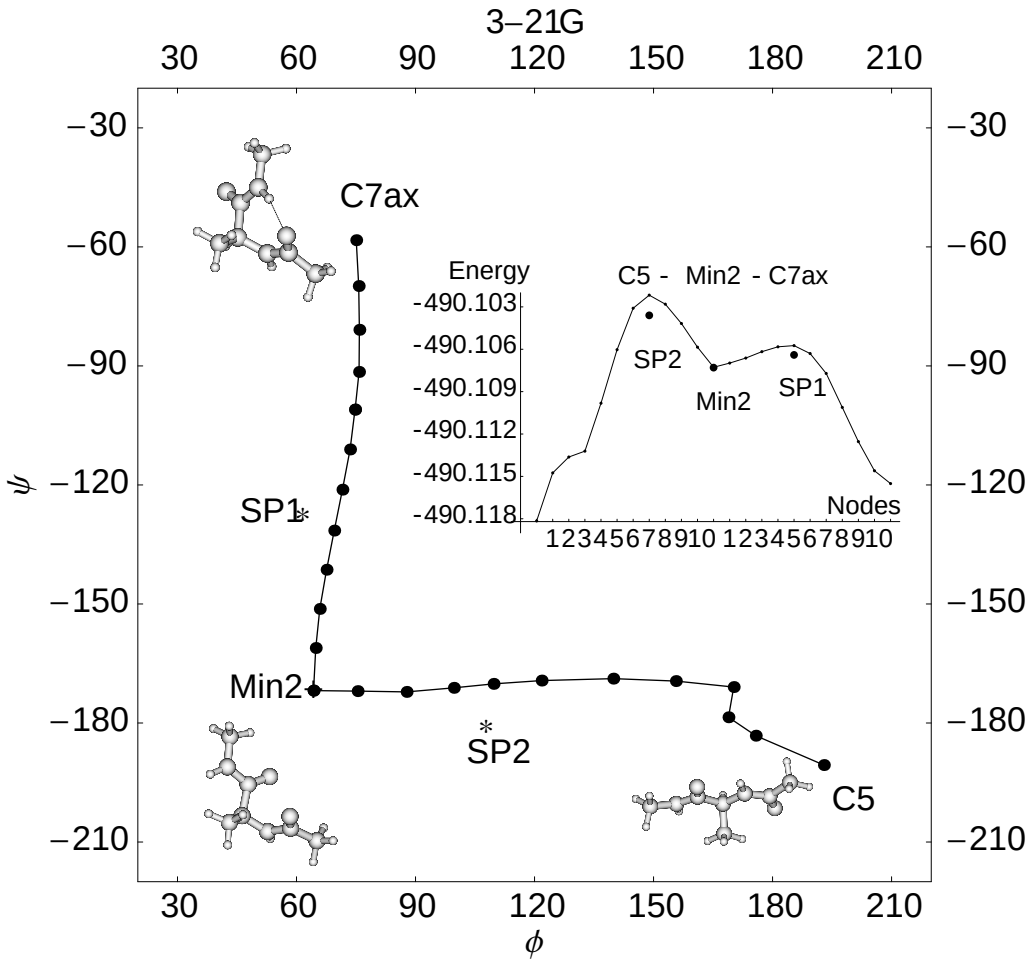


FIG. 5. Approximation of two NTs for alanine dipeptide between the C5 minimum and intermediate minimum, Min_2, and from there to the minimum $\mathrm{C7}_{\mathrm{ax}}$, as well. Method: modified CG+ optimization. Two coordinates (Φ , Ψ) are shown in a Ramachandran diagram. 10 nodes are used for every string, and thresholds ϵ=0.005, and 0.003 33, respectively. The inlay shows the energy profile of both paths in the order of their calculation.

IX. IMPLEMENTATION OF CG+ AND GAUSSIAN03 FOR THE GS METHOD

•We use scripts to call the GS-NT method in modular form including the CG+ corrector, and to call energy and gradient in an independent program: the Gaussian03.[12]

• We need the *input* file for every current geometry. The file *oben* contain the gaussian command line for the used *input* file:

Head of Gaussian03 input for corrector steps (file *oben*).

```
$RunGauss
# scf=direct 3-21G opt optcyc=1 test
```

which are followed by the actual geometry of the molecule in Cartesian coordinates; it is given by file *AlaPointG*. The Gaussian03 calls *input* and gives *output*. From the use of the Gaussian03 we need energy and gradient of the current point which we extract from file *output*; it is done by some commands of the script.

Shell script

```
call  startGS: initialization
cat   oben AlaPointG > input
call  gaussian03
grep  'SCF Done:  E(RHF) ='  output > energy
tr  -s ' ' < energy > entr
cut -d' ' -f6 entr  > energy
grep 'SCF Done:  E(RHF) =' output > ALAenergy
grep D13 output | grep estimate > ALAd13
grep D27 output | grep estimate > ALAd27
call  projector
call  predictor
#                    GS cycle
export status=0
while (test $status -ne 10) do
cat   oben AlaPointG > input
call  gaussian03
grep  'SCF Done:  E(RHF) =' output > energy
tr   -s ' ' < energy > entr
cut  -d' ' -f6 entr  > energy
grep -n 'Axes restored to original set' output | head -n 1 >noline
cut  -d':' -f1 noline > num
head -n $num output |tail -22 >gradfile
call   corrector: new GS point and test
export status=$?
if (test $status -eq 1)
then
grep 'SCF Done:  E(RHF) =' output >> ALAenergy
grep D13 output | grep estimate >> ALAd13
grep D27 output | grep estimate >> ALAd27
call  projector
call  predictor
export status=$?
fi
done
#                end GS cycle
exit
```

The procedures can be downloaded.[13]

REFERENCES

[1] W. Quapp, J. Comput. Chem. **25**, 1277 (2004).

[2] B. Peters, A. Heyden, A.T. Bell and A. Chakraborty, J. Chem. Phys. **120**, 7877 (2004).

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

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

[5] W. Quapp, M. Hirsch and D. Heidrich, Theoret. Chem. Acc. **100**, 285 (1998).

[6] W. Quapp, J. Comput. Chem. **22**, 537 (2001).

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

[8] J.C. Gilbert and J. Nocedal, SIAM J. Optimiz. **2**, 1 (1992);
G. Liu, J. Nocedal and R. Waltz, CG+ Software
www.ece.northwestern.edu/∼nocedal/CG+.html .

[9] J.J. Moré and D.J. Thuente ACM Trans. Mathem. Software **20**, 286 (1994).

[10] K. Müller and L. D. Brown, Theor. Chim. Acta **53**, 75 (1979).

[11] D.J. Wales and J.P.K. Doye, J. Phys. Chem. A **101**, 5111 (1997);
http://www-doye.ch.cam.ac.uk/jon/structures/LJ .

[12] M.J.Frisch et al., Gaussian 03, Revision B.03, Gaussian, Inc., Pittsburgh PA (2003).

[13] W. Quapp, www.math.uni-leipzig.de/∼quapp/GS

[14] M. Hirsch and W. Quapp, J. Mol. Struct. (Theochem) **683**, 1 (2004).

[15] W. Quapp, J. Theor. Comput. Chem. **2**, 385 (2003).

[16] M. Hirsch, *Zum Reaktionswegcharakter von Newtontrajektorien*, Dissertation
(Fakultät für Chemie und Mineralogie, Universität Leipzig, Dezember 2003).