JCTC Journal of Chemical Theory and Computation

Gradient Projection Method for Constraint Optimization and Relaxed Energy Paths on Conical Intersection Spaces and Potential Energy Surfaces

Bernhard Dick*

Institut für Physikalische and Theoretische Chemie, Universität Regensburg, D 93040 Regensburg, Germany

Received July 29, 2008

Abstract: A gradient projection algorithm is presented that permits the application of several constraints during geometry optimization on electronic potential energy surfaces (PES) or conical intersection (CI) seams. The algorithm generalizes the idea recently published in this journal (Sicilia et al. *J. Chem. Theory Comput.* **2008**, *4*, 257) for the optimization of conical intersection geometries. Singular value decomposition is used to transform all constraints, including those related to maintaining the CI, to a new set of constraints with orthogonal gradients. The constraints need not be satisfied at the initial geometry but will be upon convergence. A procedure is presented that determines relaxed energy paths (REP) connecting two reference structures on a potential energy surface, or the conical intersection space, without the need to assign an internal coordinate as the reaction coordinate. Examples are presented of optimizations of minimum energy structures and REPs in the CI space and REPs on a single electronic PES.

1. Introduction

During the last two decades it has been increasingly recognized that conical intersections (CIs) play a decisive role in many photochemical reactions.[1] This has also led to much interest in calculating the geometries of critical points within the conical intersection subspace. The algorithms developed so far can be roughly grouped into three families: Techniques that minimize a Lagrangian including the constraints to maintain the degeneracy of the crossing states,[2−8] gradient projection techniques,[9−14] and a method using a penalty function.[15] Recently, a comparison of these three techniques has been performed employing a semiempirical Hamiltonian.[16] This study found that the Lagrange-Newton technique needed the smallest number of iterations to converge to the minimum energy point of the conical intersection. The gradient projection technique needed usually more iterations but converged to the same result, whereas the penalty function method frequently converged to a slightly different geometry and energy. Based on the convergence performance, the Lagrange-Newton method should be the first choice. However, as discussed in refs 16 and 17, Newton−Raphson programs usually employed in quantum mechanical programs cannot be easily adapted to this technique, whereas the modifications required by the projection technique are straightforward. Also, the Hessian of the Langrangian function has negative eigenvalues for each constraint. Hence the popular Broyden-Fletcher-Goldfarb-Shanno (BFGS) update formula[18] for positive definite Hessian matrices cannot be used.

Recently, Sicilia et al. published a modified version[17] of the gradient projection technique that showed greatly improved convergence properties compared to the original technique of Bearpark et al.[9] (which was the one used in the comparative study mentioned above).[16] This modified method was also used in combination with a projection technique[19] that imposes a geometry constraint by using only that part of the search direction which is perpendicular to the gradient of the constraint. As a consequence the constraint will be fixed at the initial value. In this way points on a relaxed energy path (REP) within the conical intersection space were calculated. Apparently the method by Sicilia et al.[17] treats the constraints associated with maintaining the degeneracy between two electronic states differently from

* Corresponding author e-mail: Bernhard.Dick@chemie.uni-regensburg.de.

10.1021/ct8003029 CCC: $40.75
Published on Web 12/08/2008

an additional geometrical constraint. Care has to be taken to avoid cancelation errors that can arise when the gradient vector of the constraint is not orthogonal to the branching space.[19]

Here we present a generalization of the principle idea of ref 17 with the following properties:

1) All constraints, including those pertaining to the conical intersection, are treated in the same fashion. The algorithm can also be used for constrained optimization on a single electronic state, or for a triple degeneracy.

2) Several constraints on structural parameters (bond lengths, bond angles, or dihedral angles) or other functions of geometry (e.g., moments of inertia) can be specified at the same time.

3) A constraint can be any linear or nonlinear function of geometrical parameters for which the gradient can be calculated. For example, the projection of the geometry onto the difference vector between two reference structures can be fixed. This will lead to the REP between these two structures without the need to identify a prominent internal coordinate as the reaction coordinate.

4) The algorithm avoids problems associated with the fact that the projected Hessian does not have full rank.

5) The constraints need not be satisfied by the initial geometry but will be met upon convergence.

2. Outline of the Constrained Optimization Method

We begin with a brief review of the recent algorithm of Sicilia et al.[17] The aim is to minimize the average energy of two electronic states

$$F = \frac{1}{2}(E_1 + E_2) \tag{1}$$

subject to the two constraints

$$C^{(1)} = (E_1 - E_2) = 0 \tag{2}$$

$$C^{(2)} = H_{12} = 0 \tag{3}$$

where H_{12} is the interstate coupling matrix element. Both energies and the interstate coupling matrix element are functions of a set of N geometrical variables $\mathbf{r} = \{x_1, x_2, \cdots, x_N\}$. The original gradient projection method of Bearpark et al.[9] makes steps along the direction of a modified gradient

$$\mathbf{g}_M = \mathbf{P}\mathbf{g}_0 + \frac{2(E_1 - E_2)}{|\mathbf{g}^{(1)}|}\mathbf{g}^{(1)} \tag{4}$$

In this expression, $\mathbf{g}_0$ is the gradient of F with respect to $\mathbf{r}$, $\mathbf{g}^{(j)}$ is the gradient of the constraint $C^{(j)}$, and $\mathbf{P}$ is a projection operator that projects the component of $\mathbf{g}_0$ perpendicular to the plane spanned by $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$.

The recent improvement of this technique makes use of the Hessian matrix $\mathbf{H}$ of the problem. The algorithm switches between two modes, depending on the size of the energy difference. If this is larger than a given threshold, the gradient of eq 4 is used to calculate the step $\mathbf{d}$ according to

$$\mathbf{d} = -\mathbf{H}^{-1}\mathbf{g}_M \tag{5}$$

When the energy difference is below the threshold, the modified Hessian

$$\mathbf{H}_M = \mathbf{PHP} + (1 - \mathbf{P})\mathbf{A}(1 - \mathbf{P}) \tag{6}$$

is used, where $\mathbf{A}$ is a diagonal matrix with large diagonal elements. The procedure remedies the fact that the projected Hessian $\mathbf{PHP}$ has rank $N - 2$ and can hence not be inverted. The step is given by

$$\mathbf{d} = -\mathbf{H}_M^{-1}(\mathbf{P}\mathbf{g}_0) + \frac{E_1 - E_2}{|\mathbf{g}^{(1)}|^2}\mathbf{g}^{(1)} \tag{7}$$

In summary: above the threshold, the gradient of the energy difference is added to the projected gradient of the function F to be optimized, and the full Hessian is used. Below the threshold, the projected and rank-corrected Hessian is used, and a term proportional to the gradient of the energy difference is added to the step. Only $E_1 - E_2$ and $\mathbf{g}^{(1)}$ are used in eqs 4 and 7, i.e., the step or gradient are not modified for nonvanishing H_{12} or $\mathbf{g}^{(2)}$.

When a further constraint is added to these equations, care must be taken to ensure that application of the constraint leaves the projected gradient of the optimized function $\mathbf{P}\mathbf{g}_0$ orthogonal to the gradient of the energy difference, $\mathbf{g}^{(1)}$. This is important to ensure that upon minimization of the gradient $\mathbf{g}_M$ not only the sum of both terms goes to zero but also each term separately.[19] When the various gradient vectors $\mathbf{g}^{(k)}$ are not orthogonal, their corresponding projection operators

$$\mathbf{P}^{(k)} = 1 - \frac{\mathbf{g}^{(k)}\mathbf{g}^{(k)T}}{|\mathbf{g}^{(k)}|^2} = 1 - \mathbf{e}^{(k)}\mathbf{e}^{(k)T} \tag{8}$$

do not commute. (The superscript T indicates the transposed vector or matrix, and $\mathbf{e}$ indicates a unit vector.) It was concluded in ref 19 that the direction of the gradient of the additional constraint must be projected from $\mathbf{g}_0$, $\mathbf{g}^{(1)}$, and $\mathbf{g}^{(2)}$ prior to forming the projected gradient $\mathbf{P}\boldsymbol{g}_0$ of eq 4. Apparently the result will depend on the order in which projections are applied when several constraints are imposed and the corresponding gradients are not orthogonal.

Our generalization of the algorithm, which will be described next, avoids this problem. Our method is applicable to a generalized form of eq 1 for the definition of the optimized function, e.g.

$$F = \frac{1}{m}\sum_{j=1}^{m} E_j \tag{9}$$

With $m = 1$ a single electronic state is optimized, whereas $m = 2$ or $m = 3$ correspond to a double or triple degeneracy. This involves $(m - 1)$ constraints for the energy differences and $m(m - 1)/2$ constraints for the coupling matrix elements between all of these states. In addition to these, further constraints can be applied, e.g. on internal coordinates (bond lengths, bond angles, dihedral angles) or more generalized coordinates as discussed in the next section. We define a matrix $\mathbf{G}$ whose columns are given by the gradient vectors of all constraints that should be satisfied

$$\mathbf{G} = \{\mathbf{g}^{(1)}, \mathbf{g}^{(2)}, \cdots, \mathbf{g}^{(L)}\} \tag{10}$$

$$G_{jk} = \frac{\partial}{\partial x_j} C^{(k)} \tag{11}$$

The first two are those given in eq 2; the other can be constrained internal coordinates or other functions of the coordinates of the molecule. The dimension of this matrix is $N \times L$, where N is the number of geometrical variables, and L is the number of constraints. We assume that the number of constraints is smaller than the number of variables, i.e. $L < N$. Next this matrix is decomposed by singular value decomposition (SVD) according to

$$\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{12}$$

This decomposition is always possible, and efficient numerical algorithms exist. The matrices on the RHS of eq 12 have the following properties: Matrix **S** is a diagonal $L \times L$ matrix with positive elements. In the following we assume that they are arranged in descending order. Matrix **U** is a $N \times L$ matrix whose columns are normalized and orthogonal to each other. Finally, matrix **V** is an orthogonal $L \times L$ matrix, i.e.

$$\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = 1 \tag{13}$$

The column vectors of **U** thus form an orthonormal basis for the set of gradient vectors $\mathbf{g}^{(k)}$. The number of diagonal elements for which the ratio S_m/S_1 is above a certain threshold can be considered as the effective rank of **G**, and all basis vectors $\mathbf{u}^{(k)}$ for which this ratio is smaller than the threshold are neglected.[20] For this (potentially reduced) set of vectors the matrix **S** can be inverted, and eq 12 can be rewritten as

$$\mathbf{U} = \mathbf{G}\mathbf{V}\mathbf{S}^{-1} \tag{14}$$

Next we define linear combinations of the original constraints as

$$B^{(m)} = \sum_{j=1}^{L} C^{(j)} V_{jm} \tag{15}$$

Their gradients are given by

$$\frac{\partial}{\partial x_k} B^{(m)} = \sum_{j=1}^{L} \frac{\partial}{\partial x_k} C^{(j)} V_{jm} = \sum_{j=1}^{L} G_{kj} V_{jm} = S_m U_{km} \tag{16}$$

Thus the column vector $\mathbf{u}^{(m)}$ gives the direction of the gradient of $B^{(m)}$, and S_m is its length. In summary, the set of nonorthogonal gradient unit vectors $\mathbf{e}^{(j)} = \mathbf{g}^{(j)}/|\mathbf{g}^{(j)}|$, gradient lengths $|\mathbf{g}^{(j)}|$, and constraints $C^{(j)}$ has been transformed to an orthogonal set $\{\mathbf{u}^{(m)}, S_m, B^{(m)}\}$ from which linear dependencies have been removed.

The set of vectors $\mathbf{u}^{(m)}$ is now split into two: Set 1 contains those vectors for which $B^{(m)}/S_m$ is below a given threshold, and set 2 contains the remaining vectors. Also, two projection operators are defined

$$\mathbf{P} = 1 - \mathbf{U}\mathbf{U}^T \tag{17}$$

$$\mathbf{P}^{(1)} = 1 - \mathbf{U}^{(1)}\mathbf{U}^{(1)T} \tag{18}$$

Thus **P** projects onto the orthogonal complement of the subspace of all constraints, whereas $\mathbf{P}^{(1)}$ projects out only the components of set 1. In analogy to the method of ref 17, the gradient of the optimized function is now modified by first projecting out all contributions of the constraints and then adding contributions for those constraints that are above threshold

$$\mathbf{g}_S = \mathbf{P}\mathbf{g}_0 + 2\sum_{j}^{set2} B^{(j)}\mathbf{u}^{(j)} \tag{19}$$

It should be noted that the additional contribution to the gradient would also arise if the square of the constrained $(B^{(j)})^2$ should be minimized. Except for the first iteration, the unprojected Hessian **H** is updated at this point by the BFGS formula.[17,18] The next step taken by the algorithm is calculated from a modified Hessian without the contribution of those constraints which are below threshold

$$\mathbf{H}_S = \mathbf{P}^{(1)}\mathbf{H}\mathbf{P}^{(1)} \tag{20}$$

Note that no term was added like in eq 6 which makes the projected Hessian again full rank. This is not needed since instead of inverting the matrix the equation system

$$\mathbf{H}_S\mathbf{d}_S = -\mathbf{g}_S \tag{21}$$

is solved by a linear least-squares algorithm. This algorithm finds the shortest vector $\mathbf{d}_S$ that minimizes the sum of squares $|\mathbf{H}_S\mathbf{d}_S + \mathbf{g}_S|^2$. Finally, the step for the next iteration is calculated by adding the contributions of the constraints in set 1 as

$$\mathbf{d} = \mathbf{d}_S - \sum_{j}^{set1} \frac{B^{(j)}}{S_j}\mathbf{u}^{(j)} \tag{22}$$

3. Relaxed Energy Paths

Relaxed energy paths (REP) connecting two critical points are frequently calculated with the technique of a relaxed scan. For this purpose, an internal coordinate of the system is identified that changes monotonously and by a considerable amount along the path between two reference structures. The REP is then obtained by a sequence of constrained optimizations for a set of fixed values of this internal coordinate.

The gradient projection method outlined above is easily employed for this purpose. Since the constrained variable need not have the intended value at the beginning, each point of the REP can be started from the previously optimized point. This can be done for a single electronic potential energy surface or for the intersection space of two states.

Such a scan of one internal coordinate combined with optimization of all other coordinates requires that the projection of the REP onto the internal coordinate varies monotonously along the REP. This is not always the case, and an example will be presented in the Discussion section. Sometimes a reaction involves the collective change of several internal coordinates none of which makes a dominant contribution to the REP. Hence the possibility of finding a more general approach was considered. In particular we are interested in a strategy that connects two given points in configuration space by a REP.

An ideal algorithm would automatically identify a geometrical constraint, given by a single parameter α, such that the optimized structure for $\alpha = 0$ corresponds to the initial structure $\mathbf{r}_1$ of the REP. Optimizing the structure at $\alpha = 1$ should lead to the final structure $\mathbf{r}_2$ of the REP. Here two alternatives are presented that appear suitable for this purpose.

The first fixes the projection of the structure onto the distance vector $\Delta\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ between the initial and the final point of the REP. The constraint to be minimized is then

$$C = \frac{(\mathbf{r} - \mathbf{r}_1)\Delta\mathbf{r}}{|\Delta\mathbf{r}|^2} - \alpha \tag{23}$$

This constraint involves the coordinates of the system in absolute space. As will be shown in the application section, this formula was successful in several cases but requires that rotation and translation of the whole system is suppressed. This can be simply achieved by additional constraints but can altogether be avoided by using a constraint that makes use only of internal coordinates of the system. Our second proposal for a constraint is of this kind.

$$C = \frac{1}{N_b}\sum_{\{ij\}} \frac{d_{ij} - d_{ij}^{(1)}}{d_{ij}^{(2)} - d_{ij}^{(1)}} - \beta \tag{24}$$

In this expression, $d_{ij}^{(k)}$ is the distance between atoms i and j in structure k. The summation is performed over all bonds with an absolute change $|d_{ij}^{(2)} - d_{ij}^{(1)}|$ above a threshold, which is usually set as a certain fraction of the largest change. N_b is the number of these bonds. This formula is especially useful in cases where the main contribution of the reaction coordinate shifts from one internal coordinate to another one along the relaxed scan. Equation 24 can easily be extended to include other internal coordinates like bond angles or dihedral angles.

4. Computational Details

The algorithm was coded into a FORTRAN program that acts as an interface to the quantum chemical program GAMESS[21] optimized for the Windows operation system.[22] It generates the appropriate input files, starts GAMESS, and extracts the required energies and energy gradients from the GAMESS output files. The constraints and their gradients are calculated in the external program. Singular value decomposition is performed by the routine DGESDD from the LaPack library,[23] the linear least-squares problem of eq 20 is solved by the routine DGELSD from the same library.

The program always operates on the mass- and symmetry-weighted unique Cartesian coordinates of the system. When the norm of the calculated geometry step was larger than a given value d_{max}, the step was rescaled accordingly. Iteration was stopped when the norm of the gradient vector and the step were both below a certain threshold, which was set to 0.003 for all calculations in this work. The threshold on the condition number of the gradient matrix was set such that the vectors $\mathbf{u}^{(m)}$ with $S_m/S_1 < 10^{-6}$ were neglected. The threshold for including a vector in set 1 was $B^{(m)}/S_m < 5 \times 10^{-3}$. With this setting the constraints were satisfied to within $10^{-6}-10^{-8}$ units, and energies changed by less than 10^{-6} Hartree during the last few steps.

5. Results and Discussion

5.1. Convergence. We have chosen the CI between S_0 and S_1 of fulvene as an example in order to permit comparison with the work of Bearpark et al.[19] The first

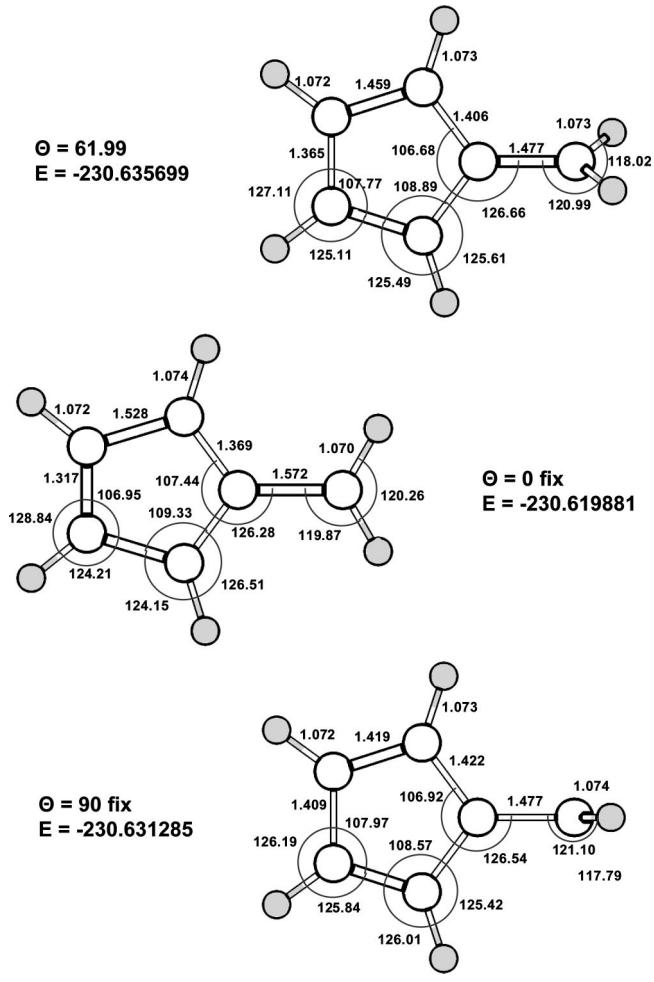

Figure 1. Optimized bond lengths (Angstrom units) and angles (in degrees) for the minimum energy point of the S_0/S_1 conical intersection in fulvene without geometrical constraint (top), with fixed dihedral angle $\theta = 0°$ (middle), and with fixed dihedral angle $\theta = 90°$ (bottom).

geometry of this CI reported by Dreyer and Klessinger[35] has the exocyclic CH_2 group oriented perpendicular to the ring (CI_{Perp}). Bearpark et al.[32] arrived at the same conclusion but showed in addition that the intersection exists for all twist angles up to the planar structure. Later it was recognized that a C_2-symmetric structure with a twist angle of $\approx 63°$ (CI_{63}) is ≈ 2.3 kcal mol^{-1} lower in energy.[34] The relaxed energy path connecting the planar and the perpendicular structures on the seam has been presented in ref 19, and it is this calculation that we take as the benchmark here. It should, however, be mentioned that recently a C_1-symmetric structure has been presented[33] that is still 0.003 kcal mol^{-1} lower in energy than CI_{63}.

In our calculation, the same CASSCF(6,6)/6−31G* type of wave function and C_2 symmetry as in ref 19 was employed. Optimizations were performed for the minimum energy point of the CI without constraint and with the dihedral angle between the five-membered ring and the CH_2 group constrained to $\theta = 0°$ and $\theta = 90°$, respectively. All optimizations were started from the geometry of the electronic ground-state optimized at the RHF/6−31G* level.

The optimized bond lengths and angles of the three structures are shown in Figure 1. They compare well with

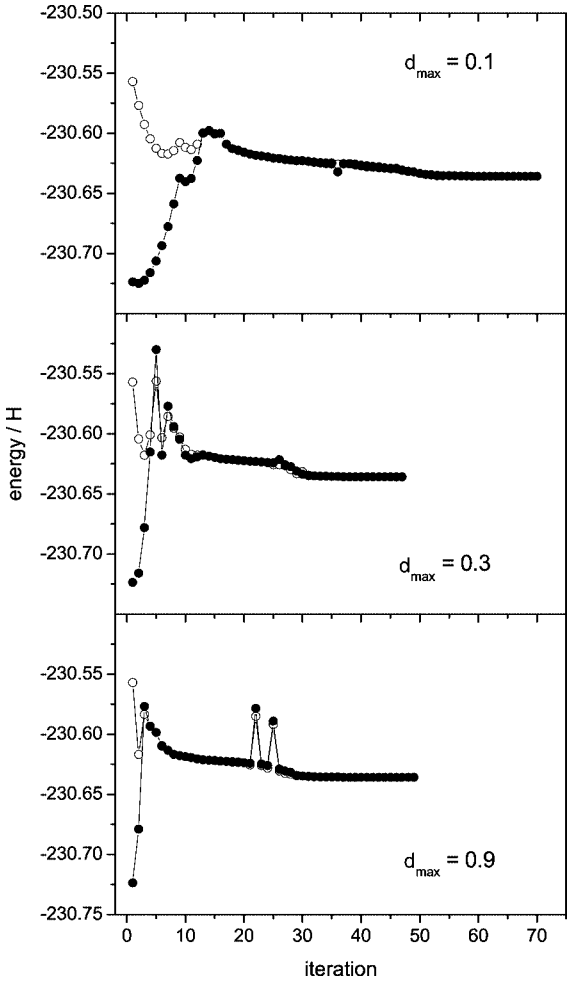

Figure 2. Convergence behavior of the algorithm for optimization of the S_0/S_1 conical intersection in fulvene without geometrical constraint. Three different values for the norm of the largest step d_{max} were used. Full and open symbols represent ground-state and excited-state energies, respectively.

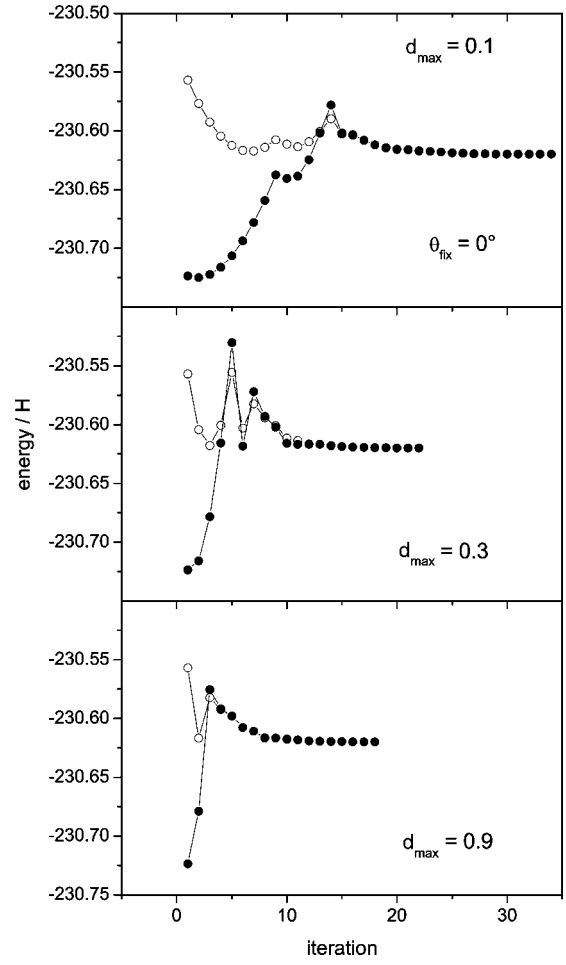

Figure 3. Convergence behavior of the algorithm for optimization of the S_0/S_1 conical intersection in fulvene with fixed dihedral angle $\theta = 0°$. Three different values for the norm of the largest step d_{max} were used. Full and open symbols represent ground-state and excited-state energies, respectively.

those reported by Bearpark et al.[19] However, we perform separate state specific CASSCF calculations for the S_0 and the S_1 states, which belong to different irreducible representations of the C_2 point group. This ensures orthogonality of the two wave functions but allows for different optimized orbitals. As a consequence, our energies are slightly lower than those reported in ref 19.

Each optimization was repeated three times with the setting of the maximum allowed step size d_{max} successively increased by a factor of 3, beginning with $d_{max} = 0.1$ Å. The observed convergence behavior is shown in Figures 2−4.

Without geometrical constraint and with the smallest allowed step size $d_{max} = 0.1$ Å (upper panel in Figure 2), the degeneracy is found after 12 steps, with the energy difference below 1 mH from this point on. Convergence requires 58 more steps during which the energy of both states relaxes by 36 mH. When the smallest allowed step size is increased to $d_{max} = 0.3$ Å (middle panel in Figure 2), the degeneracy is again found after 12 steps within 1 mH. However, the energies show large oscillations during the first few steps, which can even bring the A-state considerably above the B-state. Convergence is reached after a total of 47 steps, with almost the same energy (6 μH lower) than in the previous run. Interestingly, with an even larger value of $d_{max} = 0.9$ Å (lower panel in Figure 2) convergence to the degeneracy is already reached after 3 steps, and the initial oscillations disappear. The final energy after 48 steps is 2 μH lower than in the previous case, and the splitting of the states amounts to 6×10^{-8} H.

A similar behavior is observed for the optimization with the dihedral angle $\theta = 0°$ fixed (Figure 3). Convergence is slow but smooth with $d_{max} = 0.1$ Å, it becomes faster but shows initial oscillations with $d_{max} = 0.3$ Å, and performs best with $d_{max} = 0.9$ Å, requiring only 17 steps. The fast convergence is mainly due to the fact that the fixed dihedral angle had initially already the required value.

However, the algorithms performs also well when the dihedral angle is initially far away from the target value, as seen in Figure 4 for the target $\theta = 90°$. With a small step size the simultaneous optimization toward $\Delta E = 0$ and $\theta = 90°$ apparently leads into a structural region where the energy is much too high, but the algorithm finally finds its way down to the optimized perpendicular structure of fulvene. The escape in that unfavorable region is much shorter when larger steps are allowed ($d_{max} = 0.3$ Å, middle panel in Figure 4),

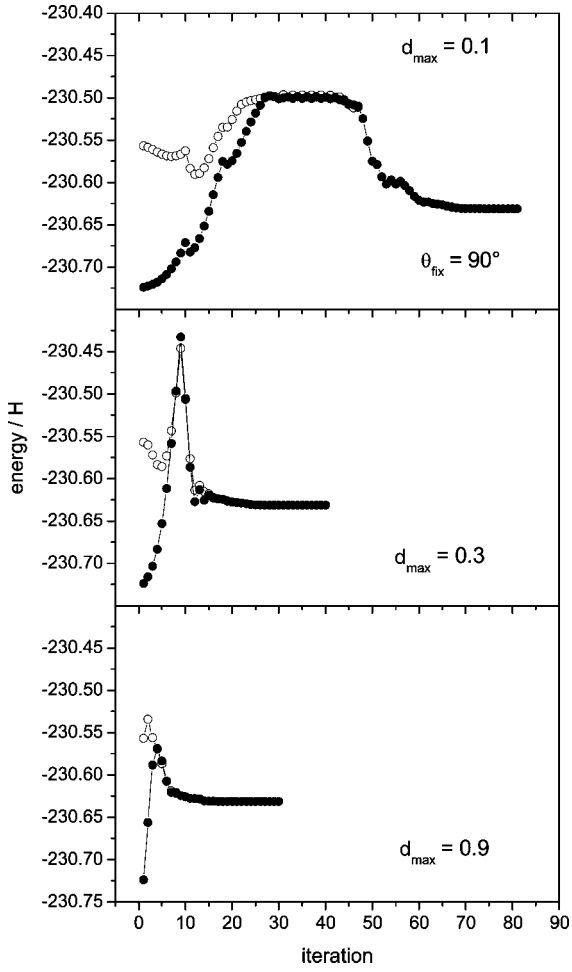

Figure 4. Convergence behavior of the algorithm for optimization of the S_0/S_1 conical intersection in fulvene with fixed dihedral angle $\theta = 90°$. Three different values for the norm of the largest step d_{max} were used. Full and open symbols represent ground-state and excited-state energies, respectively.

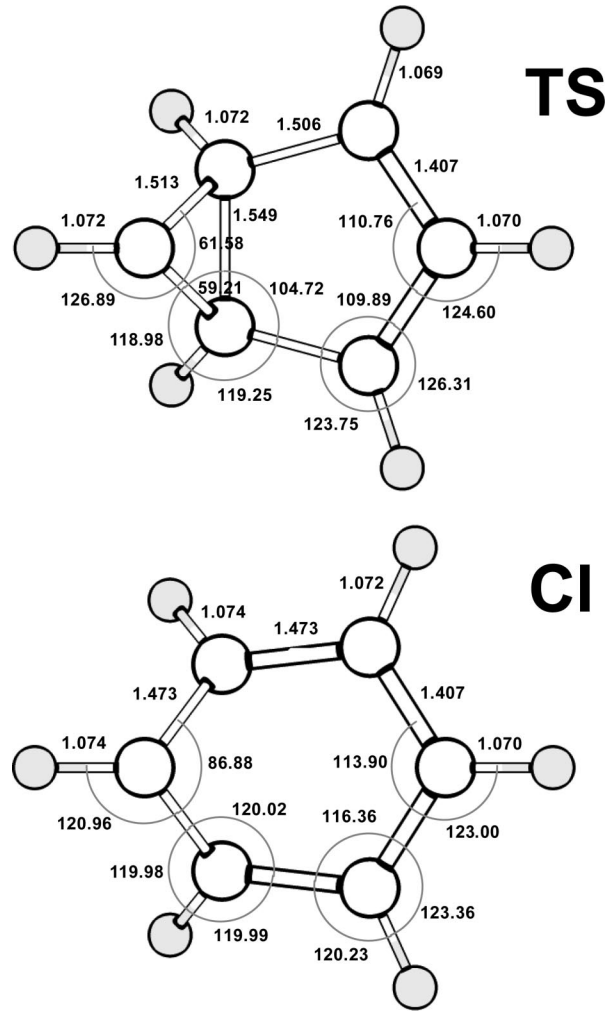

Figure 5. Structures of the optimized S_0/S_1 conical intersection (CI) of benzene and the transition state (TS) between two benzvalene isomers. Bond lengths are given in Angstroms, bond angles in degrees.

and convergence is again fastest with the largest allowed step size of $d_{max} = 0.9$ Å (lower panel in Figure 4).

As a common feature of all these convergence studies one observes that convergence is smooth once the degeneracy has been reached to within a few mH. Only in one case (lower panel of Figure 2) two short excursions (one step each) occurred into an unfavorable region of the PES before the system continued to convergence. In this case very large steps (0.9 Å) were permitted. It might be that in this case the Hessian was not yet well balanced after only 20 updates. In the early phase of the optimization where the two states are widely apart, oscillations and excursions in highly excited regions are more frequent, especially when large step sizes are permitted. Hence one can expect even better convergence behavior when the maximum allowed step size is dynamically adjusted according to the splitting of the states and the success of the previous steps taken by the algorithm.

The algorithm does not always go easily to the conical intersection of interest when started from a point far away. E.g., the optimization of the S_0/S_1 conical intersection of benzene seems to depend strongly on the precise structure used as the starting point. Using different distortions from the ground-state structure as starting structures our algorithm optimized to various local minima of the CI. This might be improved by using a more sophisticated guess for the initial Hessian (in this paper a diagonal matrix is used throughout) or dynamically adjusting the trust radius for the Newton-Raphson step. Alternatively, the degeneracy of interest can be located quickly by an algorithm based on the idea of Longuet-Higgins loops.[24] For the discussion of the underlying principles see refs 25−29. As an example a CASSCF(8,7)/DZV calculation of the S_0/S_1 conical intersection of benzene is shown below.[30] This CI is encircled by a loop that has benzene and two of its benzvalene isomers as its anchors.[26] The structure of the transition state between the two benzvalene isomers is shown in Figure 5. It has A″ symmetry in the C_S point group. Classifying the states of benzene with reference to the same mirror plane, the ground-state has A′ symmetry. Along a path between these two structures that conserves the mirror plane the two states must hence cross. This crossing is located within three steps by a bisection algorithm as shown in the upper part of Figure 6. Since this algorithm does not require gradients, each iteration is even faster than those of the optimization algorithm. Starting from this structure the present algorithm smoothly converges to

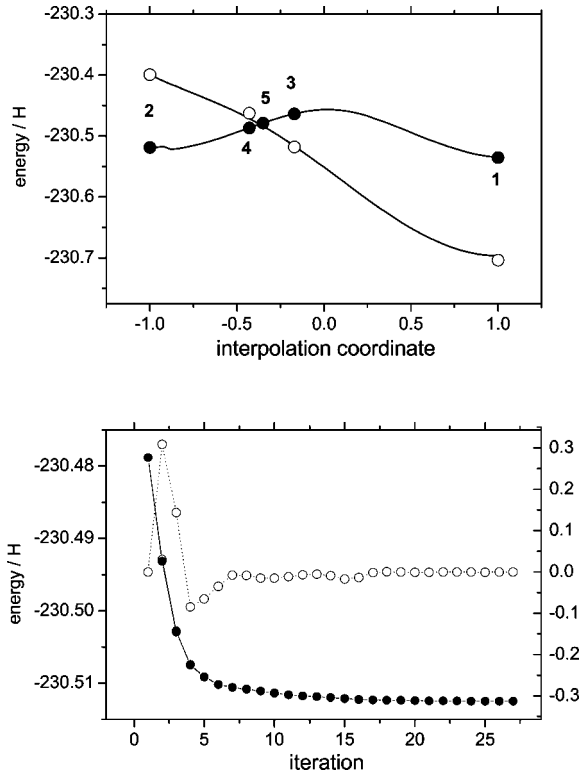

Figure 6. Upper part: Search for the crossing seam along the coordinate defined as the linear interpolation between the structures of benzene (1) and the transition state between two benzvalene isomers (2). The points (3-5) indicate the steps taken by the search algorithm to locate the degeneracy. Open symbols represent the A′ state, full symbols the A″ state. The full lines represent the two electronic potential energy curves. Lower part: Convergence of the optimization of the CI. Full symbols: energy (in Hartree units, left scale) of the degenerate pair of states, open symbols: Splitting of the states (in milli-Hartree units, right scale).

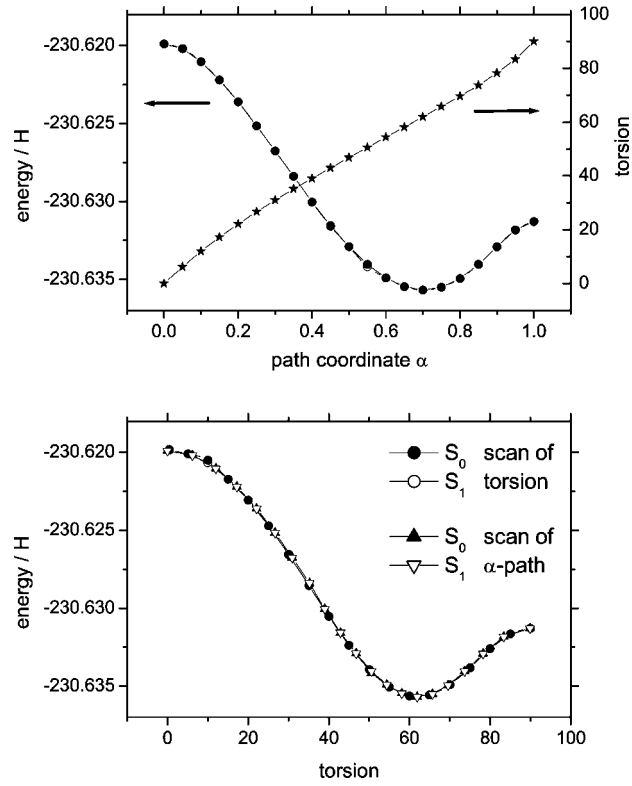

Figure 7. REP connecting the planar and the perpendicular structures of fulvene on the S_0/S_1 conical intersection seam. Upper panel: Energies (left scale) and torsional angle θ (right scale) of the optimized REP along the path coordinate α (eq 23). Lower panel: Energies of the REP as function of the torsional angle. Circles are optimized at equidistant points along θ, and triangles are the results of the path calculated along the path coordinate α defined in eq 23.

the optimized structure of the CI while maintaining the degeneracy within less than 0.3 mH, as seen in the lower part of Figure 6. The structural parameters of the optimized CI are given in Figure 5. They compare well with those presented in the first report on this CI by Palmer at al.[31]

5.2. Relaxed Energy Paths. As first example we present calculations of the REP connecting the planar and perpendicular transition states on the intersection space for the CI between S_0 and S_1 of fulvene. The same basis and active space as in the previous section was used. The REP was calculated along the torsional coordinate θ (19 equidistant points with 5° distance) and also along the path coordinate α defined in eq 23 (21 points with distance of 0.05). Both runs were started from the optimized planar CI geometry and required a total of 107 function evaluations for all points to converge. The results are displayed in Figure 7. The upper panel shows the energy values of the relaxed scan at fixed values of the generalized path coordinate α defined in eq 23 (left scale). The torsional angle at the optimized points increases monotonously with the path parameter α (right scale). The lower panel shows the same REP, now calculated along equidistant points with fixed torsional angle. These data are represented by circles. The data from the scan along α have been included in this plot at the appropriate torsional angles and are shown as triangles. Apparently both scans walked along the same REP.

It should be mentioned that the use of the path parameter α requires the application of further constraints since it is defined in terms of absolute coordinates in $3N$ space. Hence a given projection of the shift $\mathbf{r} - \mathbf{r}_1$ onto the difference vector $\mathbf{r}_2 - \mathbf{r}_1$ between the two reference structures could be achieved by a rotation or translation of the whole molecule. This must be suppressed by fixing components of the center of mass **R** and off-diagonal elements of the inertial tensor **I**. In the present case with C_2-symmetry these conditions were $R_z = 0$ and $I_{xy} = 0$.

As a second example the tautomerization reaction on the PES of the electronic ground-state in propandial was studied. All calculations were performed with CASSCF(10,7)/6−31G** wave functions. The active space consisted of the 5 valence π-orbitals and two σ-orbitals representing the O−H bond and the oxygen lone pair, respectively. The optimized structures of one of the two equivalent ground-state tautomers and the transition state (TS) are shown in Figure 8. The REP was first explored by performing a relaxed scan with the distance r_{19} as the parameter (for the numbering of atoms see Figure 8). The REP obtained in this way is shown by the open circles in Figure 9. Beginning with the value $r_{19} =$ 1.99 Å, the energy rises monotonously with smaller values of r_{19} and reaches a maximum near $r_{19} = 1.2$ Å, in

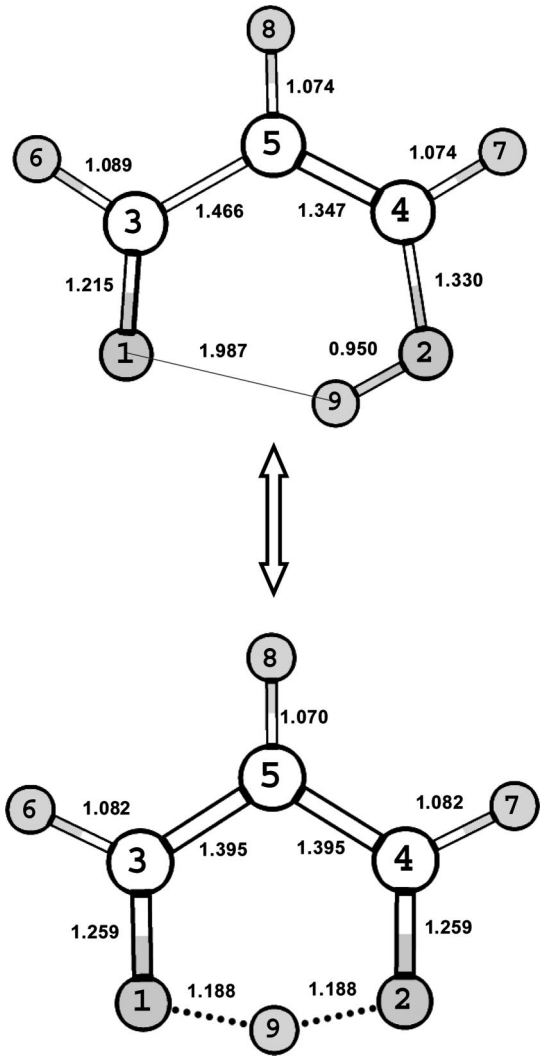

Figure 8. Optimized structures of the electronic ground state (top) and the transition state (bottom) of the enol form of propandial. Bond lengths are given in Angstroms, bond angles in degrees.

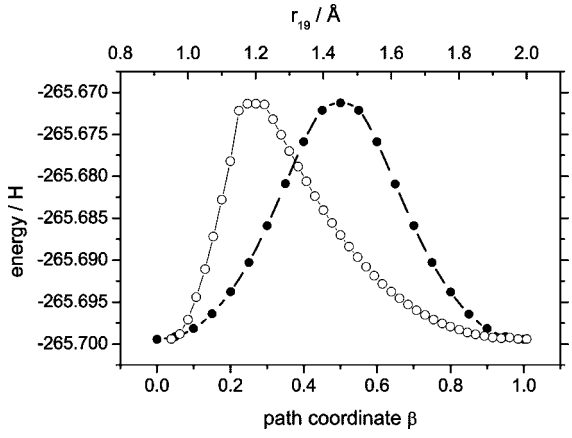

Figure 9. Energy profile on the S_0 potential energy surface connecting the two enol forms of propandial. Open circles (upper axis): relaxed scan along r_{19}. Full cirlces (lower axis): relaxed scan along the path coordinate β defined in eq 24.

accordance with the structure of the TS. The REP then decays quickly to the initial energy value, now at the other tautomer of the ground state. The REP appears unsymmetric: The

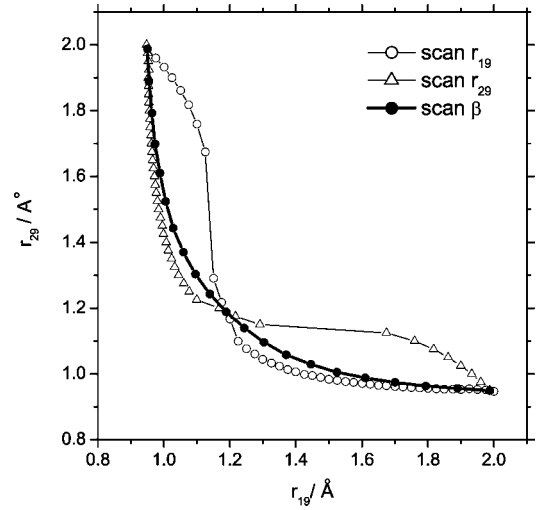

Figure 10. Projection of the REP between the two enol forms of propandial onto the plane spanned by the two distances between the migrating H-atom and the two O-atoms. Open circles: Path followed by a relaxed scan with equidistant steps in r_{19}. Open triangles: Path followed by a relaxed scan with equidistant steps in r_{29}. Full circles: Path followed by a relaxed scan with equidistant steps in the path coordinate β defined in eq 24.

change of the O−H distance amounts to 0.80 Å on one side of the TS but only 0.24 Å at the other side. Figure 10 reveals another problem with this type of relaxed scan. This figure shows the projection of the path onto the plane spanned by the two OH distances. The path mentioned above is represented by the open circles. Beginning with $r_{19} = 1.99$ Å the path shows only very little change in the other distance r_{29}, so that the reaction coordinate is well represented by the internal coordinate r_{19}. However, shortly after passing through the TS, the distance r_{29} changes by a large amount of 0.4 Å, whereas r_{19} changes by only 0.025 Å. Apparently, in this region, r_{19} has only a small projection on the reaction coordinate and is a bad choice for the parameter of a relaxed scan. In addition, the path obtained when following r_{19} is not symmetric with respect to the exchange of r_{19} and r_{29}. This becomes apparent when the data points corresponding to the scan in the opposite direction, now with equidistant steps in r_{29}, is included in the plot (represented by the open triangles). Such a symmetry is, however, expected from the fact that the two minima correspond to equivalent structures. Hence we conclude that the REP calculated as a relaxed scan along either of the two O−H distance coordinates deviates from the true REP due to the switch between the dominant contributions of the two internal coordinates to the reactions coordinate.

In this case the path coordinate β defined in eq 24 performs much better. The energy profile, shown by the full circles in Figure 9, is symmetric around $\beta = 0.5$, which corresponds to the TS structure. For each optimized point along this path the pairs of distances of the hydrogen atom to both oxygen atoms are plotted in Figure 10 as full circles. It is obvious that this path is symmetric with respect to the exchange of these two bond lengths. It is also seen that the point for $\beta = 0.5$ is at the crossing point of the two paths discussed before. Hence, all three paths go through the same TS, as is also

apparent from the same peak height of the energy profiles in Figure 9.

6. Conclusions

An algorithm has been presented for optimization of molecular structures subject to geometric and energetic constraints. The algorithm is a generalization of a recently published method developed in particular for the optimization of conical intersections. The two algorithms are identical for the particular case that a conical intersection is optimized, the interstate coupling gradient vanishes (e.g., due to symmetry), and no additional geometrical constraints are used. The algorithm presented here treats all constraints in an equivalent way. Hence a term is added to the projected gradient (eq 19) or the search step (eq 22) when the interstate coupling matrix element or its gradient does not vanish. The original algorithm adds such a term only for the energy difference. Since the interstate coupling gradient is not available in the GAMESS program, all examples used in the present study were chosen such that the interstate coupling vanishes by symmetry.

The algorithm works well with and without an additional geometry constraint, even when the starting geometry is far from satisfying the constraint. The algorithm apparently tolerates rather large step sizes, and rescaling the steps to a smaller size frequently leads to a worse performance, although the converged results are very similar. Still better convergence behavior can be expected from strategies that dynamically adjust the maximum step size depending on the performance of the previous steps.

The algorithm usually works well with a diagonal matrix for the initial Hessian matrix, but the use of a more realistic Hessian was also studied. E.g., when the Hessian of the RHF-optimized ground-state of fulvene was used to initiate the optimization of the conical intersection, a much faster convergence was found for the full optimization and the optimization with the constraint of $\theta = 0$. However, convergence was much slower in the case $\theta = 90°$. Apparently the Hessian for the planar structure is a bad guess for the perpendicular CI structure.

Since the number of variables and the size of the Hessian matrix are not modified, the algorithm should be easily incorporated into existing Newton-Raphson type optimization routines in quantum chemical programs.

Acknowledgment. This work has been supported by the Fonds der Chemischen Industrie.

References

(1) Domcke, W.; Yarkony, D. R.; Köppel, H. *Conical Intersections: Electronic Structure, Dynamics & Spectroscopy*; World Scientific Publishing Co.: Singapore, 2004.

(2) Yarkony, D. R. *J. Chem. Phys.* **1990**, *92*, 2457.

(3) Yarkony, D. R. *J. Phys. Chem. A* **2004**, *108*, 3200.

(4) Manaa, M. R.; Yarkony, D. R. *J. Chem. Phys.* **1993**, *99*, 5251.

(5) Anglada, J. M.; Bofill, J. M. *J. Comput. Chem.* **1996**, *18*, 992.

(6) De Vico, L.; Olivucci, M.; Lindh, R. *J. Chem. Theory Comput.* **2005**, *1*, 1029.

(7) Dallos, M.; Lischka, H.; Shepard, R.; Yarkony, D. R.; Szalay, P. G. *J. Chem. Phys.* **2004**, *120*, 7330.

(8) Ragazos, I. N.; Robb, M. A.; Bernardi, F.; Olivucci, M. *Chem. Phys. Lett.* **1992**, *197*, 217.

(9) Bearpark, M. J.; Robb, M. A.; Schlegel, H. B. *Chem. Phys. Lett.* **1994**, *223*, 269.

(10) Toniolo, A.; Ben-Nun, M.; Martinez, T. J. *J. Phys. Chem. A* **2002**, *106*, 4679.

(11) Chachiyo, T.; Rodriguez, J. H. *J. Chem. Phys.* **2005**, *123*, 094711.

(12) Yamazaki, S.; Kato, S. *J. Chem. Phys.* **2005**, *123*, 114510.

(13) Izzo, R.; Klessinger, M. *J. Comput. Chem.* **2000**, *21*, 52.

(14) Page, C. S.; Olivucci, M. *J. Comput. Chem.* **2003**, *24*, 298.

(15) Ciminelli, C.; Granucci, G.; Persico, M. *Chem. Eur. J.* **2004**, *10*, 2327.

(16) Keal, T. W.; Koslowski, A.; Thiel, W. *Theor. Chem. Acc.* **2007**, *118*, 837.

(17) Sicilia, F.; Blancafort, L.; Bearpark, M. J.; Robb, M. A. *J. Chem. Theory Comput.* **2008**, *4*, 257.

(18) Fletcher, R. *Practical methods of optimization*; Wiley: Chichester, 1987.

(19) Bearpark, M. J.; Blancafort, L.; Paterson, M. J. *Mol. Phys.* **2006**, *104*, 1033.

(20) A set of orthogonal vectors spanning the space of the constraints gradients can also be generated by Gram-Schmidt orthogonalization. In this case care has to be taken to detect linear dependencies. The advantage of SVD is that, in addition to providing an orthogonal space, the magnitudes of the gradient vectors is transformed to the weights S_m.

(21) (a) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347. (b) GordonM. S.; SchmidtM. W. In *Theory and Applications of Computational Chemistry, the first forty years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005.

(22) (a) Granovsky, A. A. *PC GAMESS version 7.0*. http://classic.chem.msu.su/gran/gamess/index.html (accessed Oct. 28, 2008). (b) Nemukhin, A. V.; Grigorenko, B. L.; Granovsky, A. A. *Moscow Univ. Chem. Bull.* **2004**, *45*, 75.

(23) http://www.netlib.org/lapack/ (accessed Oct. 28, 2008).

(24) Dick, B.; Haas, Y.; Zilberg, S. *Chem. Phys.* **2008**, *347*, 65.

(25) (a) Herzberg, G.; Longuet-Higgins, H. C. *Discuss. Faraday Soc.* **1963**, *35*, 77. (b) Longuet-Higgins, H. C. *Proc. Roy. Soc., London A* **1975**, *344*, 147.

(26) Zilberg, S.; Haas, Y. *Chem. Eur. J.* **1999**, *5*, 1755.

(27) (a) Zilberg, S.; Haas, Y. *Chem. Phys.* **2000**, *259*, 249. (b) Haas, Y.; Zilberg, S. *Adv. Chem. Phys.* **2002**, *124*, 433. (c) Haas, Y.; Cogan, S.; Zilberg, S *Int. J. Quantum Chem.* **2005**, *102*, 961. (d) Zilberg, S.; Haas, Y. *J. Phys. Chem. A* **2003**, *107*, 1222.

(28) The loop approach has been critisized by Vanni et al.[29] These authors conclude in particular that it will be difficult to transform that idea into a computer algorithm. A detailed comparison of the approaches presented by Zilberg and Haas with that of Vanni et al. is far beyond the scope of the present paper. Here it should suffice to note that the Zilberg/Haas

approach did work in many cases, in particular when symmetry could be used.

(29) Vanni, S.; Garavelli, M.; Robb, M. A. *Chem. Phys.* **2008**, *347*, 46.

(30) The minimum active space is CAS(6,6). The bisection algorithm makes, however, very large jumps in configuration space. E.g., the CAS for benzene is started with the orbitals of the TS structure. I observed that in these searches the MCSCF converges faster when the active space is slightly larger than minimal. In the present case one additional occupied orbital was enough.

(31) Palmer, I. J.; Ragazos, I. N.; Bernardi, F.; Olivucci, M.; Robb, M. A. *J. Am. Chem. Soc.* **1993**, *115*, 673.

(32) Bearpark, M. J.; Bernardi, F.; Olivucci, M.; Robb, M. A.; Smith, B. R. *J. Am. Chem. Soc.* **1996**, *118*, 5254.

(33) Sicilia, F.; Bearpark, M. J.; Blancafort, L.; Robb, M. A. *Theor. Chem. Acc.* **2007**, *118*, 241.

(34) Paterson, M. J.; Bearpark, M. J.; Robb, M. A.; Blancafort, L.; Worth, G. A. *J. Chem. Phys.* **2004**, *121*, 11562.

(35) Dreyer, J.; Klessinger, M. *J. Chem. Phys.* **1994**, *101*, 10655.

CT8003029