Theor Chem Acc (1998) 100:265–274
DOI 10.1007/s00214980m135

Theoretical
Chemistry Accounts

Regular article

On the automatic restricted-step rational-function-optimization method

Emili Besalú[1], **Josep Maria Bofill**[2]

[1] Institut de Química Computacional i Departament de Química, Universitat de Girona, Campus de Montilivi, E-17071 Girona, Catalunya, Spain
[2] Departament de Química Orgànica, Universitat de Barcelona, Marti i Franquès 1, E-08028 Barcelona, Catalunya, Spain

Received: 18 June 1998 / Accepted: 17 September 1998 / Published online: 23 November 1998

Abstract. The rational function optimization algorithm is one of the widely used methods to search stationary points on surfaces. However, one of the drawbacks of this method is the step reduction procedure to deal with the overstepping problem. We present and comment on a method such that the step obtained from the solution of the rational function equations possesses the desired correct length. The analysis and discussion of the method is mainly centered on the location and optimization of transition states.

Key words: Rational function optimization method – Molecular geometry optimization – Transition state geometry optimization – Restricted step algorithm – Augmented Hessian algorithm

1 Introduction

In theoretical chemical physics it is important to find stationary points (minima and first-order saddle points) on potential energy surfaces because they correspond to equilibrium and transition state structures. Many algorithms exist regarding the search for extrema on multidimensional potential energy surfaces (see e.g. [1] and references therein). One of the most powerful and efficient algorithms is that proposed by Banerjee et al. [2], the so-called rational function optimization (RFO). Basically, the algorithm consists in approximating the quadratic variation of the energy in the neighborhood of a point $\mathbf{q}$

$$\mathrm{Q}(\Delta\mathbf{q}) = \mathbf{g}^{\mathrm{T}}\Delta\mathbf{q} + \tfrac{1}{2}\Delta\mathbf{q}^{\mathrm{T}}\mathbf{H}\Delta\mathbf{q} \tag{1}$$

by a [2/2] Padé approximant

$$q(\Delta\mathbf{q}) = \frac{\mathrm{Q}(\Delta\mathbf{q})}{1+\Delta\mathbf{q}^{\mathrm{T}}\mathbf{S}\Delta\mathbf{q}} = \frac{\frac{1}{2}\begin{pmatrix}1 & \Delta\mathbf{q}^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}0 & \mathbf{g}^{\mathrm{T}}\\ \mathbf{g} & \mathbf{H}\end{pmatrix}\begin{pmatrix}1\\ \Delta\mathbf{q}\end{pmatrix}}{\begin{pmatrix}1 & \Delta\mathbf{q}^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}1 & \mathbf{0}^{\mathrm{T}}\\ \mathbf{0} & \mathbf{S}\end{pmatrix}\begin{pmatrix}1\\ \Delta\mathbf{q}\end{pmatrix}} \tag{2}$$

where the step vector $\Delta\mathbf{q}$ gives the correction of the molecular geometry, $\mathbf{g}$ is the gradient vector, and $\mathbf{H}$ the Hessian matrix of the energy at $\mathbf{q}$. The symmetric $\mathbf{S}$ matrix has to be specified but normally is taken as the unit matrix $\mathbf{I}$. The matrix that appears in the numerator of Eq. (2) is the so-called augmented Hessian (AH). As pointed out by Jensen [3], when the RFO method is applied to locate transition states, very often the step length is so large that the correction gives a geometry outside the region where the second-order expansion is valid. This region, the so-called "trust region", is characterized by a ball radius R, the "trust radius". Using the RFO algorithm, if $(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})^{1/2} > R$, one scales the vector $\Delta\mathbf{q}$ by a factor. However, in this situation $\Delta\mathbf{q}$ does not possesses the correct direction [3–5]. On the other hand, it has been demonstrated that an inappropriate determination of the optimal direction of the $\Delta\mathbf{q}$ vector within the "trust region" may produce an increase of the number of iterations or a divergence in the optimization process [6].

Obtaining vector $\Delta\mathbf{q}$ using the RFO with the appropriate step length had been considered by Banerjee et al. [2]. These authors make a dynamic step reduction by performing a one-dimensional line search along $\mathbf{q}+\alpha\Delta\mathbf{q}$, where $\alpha = \beta(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})^{-1/2}$. The β parameter is chosen in such a way that the new step length, $\alpha\Delta\mathbf{q}$, falls in the "trust region". However, the resulting scaled step vector does not satisfy the optimal solution of Eq. (2). An alternative method was given by Jensen and Jørgensen [5], consisting in taking $\mathbf{S}=\mathbf{I}$ and replacing the numerator in Eq. (2) by the following parametric form:

$$\begin{pmatrix}1 & \Delta\mathbf{q}^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}0 & \alpha\mathbf{g}^{\mathrm{T}}\\ \alpha\mathbf{g} & \mathbf{H}\end{pmatrix}\begin{pmatrix}1\\ \Delta\mathbf{q}\end{pmatrix} \tag{3}$$

where the parameter α is determined using the equation $(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})^{1/2} = R$. Note that, in this case, $\Delta\mathbf{q}$ is a function of α, e.g. $\Delta\mathbf{q}(\alpha)$. As pointed out by Khait et al. [7], the

Correspondence to: J.M. Bofill
e-mail: jmbofill@canigo.qo.ub.es

disadvantage of this approach is the lack of analytic dependence of $\Delta\mathbf{q}(\alpha)$ with respect to the α parameter. However, Jensen and Jørgensen [5] determined a value for α for which $(\Delta\mathbf{q}(\alpha)^{\mathrm{T}}\Delta\mathbf{q}(\alpha))^{1/2} \approx R$, within 10%. On the other hand, Khait et al. [7] suggest another technique which is quite close to that proposed by Banerjee et al. [2]. In this technique the step length is obtained by minimizing the absolute value of the scalar product between the gradient $\mathbf{g}(\alpha)$ and the line $\mathbf{q} + \alpha\Delta\mathbf{q}$, $\mathbf{g}(\alpha)$ is the gradient calculated at the point $\mathbf{q} + \alpha\Delta\mathbf{q}$. Finally, Anglada and Bofill [8] proposed a RFO method where the AH matrix is shifted in such a way that by solving Eq. (2) one already obtains a step vector $\Delta\mathbf{q}$ that satisfies the condition $(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})^{1/2} = R$. This method can be seen as an interplay between the RFO and the (QA/TRIM) [9a, 9b] or restricted quasi-Newton-Raphson (RQNR) [9c] algorithms.

Considering the previous revision and with the aim that the RFO solution falls into the "trust region", we propose in this article a RFO method such that its optimal solution also satisfies exactly the condition $(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})^{1/2} = R$. We study also the connection between the proposed RFO method and the QA/TRIM and RQNR techniques [9]. The new RFO technique will be applied in the location and optimization geometries of both minima and transition structures, but the discussion is mainly focused on transition states.

2 Theoretical background

2.1 Summary of the RFO algorithm

We are interested in locating stationary points (minima, maxima, saddle points of any order) of a continuous and differentiable function $E(\mathbf{q})$ depending on n unconstrained variables $\mathbf{q}^{\mathrm{T}} = (q_1, \ldots, q_n)$. In the RFO method the new point $\mathbf{q}$ is simply taken as $\mathbf{q} + \Delta\mathbf{q}$, where the step vector $\Delta\mathbf{q}$ extremalizes $q(\Delta\mathbf{q})$, given by Eq. (2), that is $\nabla q(\Delta\mathbf{q}) = \mathbf{0}$. The latter condition yields as eigenvalue equation of dimension $(n+1)$ [2, 10]:

$$\begin{pmatrix} 0 & \mathbf{g}^{\mathrm{T}} \\ \mathbf{g} & \mathbf{H} \end{pmatrix} \begin{pmatrix} v_{1,i} \\ \mathbf{v}_{n,i} \end{pmatrix} = \lambda_i \begin{pmatrix} 1 & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \begin{pmatrix} v_{1,i} \\ \mathbf{v}_{n,i} \end{pmatrix} \qquad \forall i, i = 1, \ldots, n+1 \tag{4}$$

where $\mathbf{v}_{n,i}$ is the projection of the eigenvector $\mathbf{v}_i$ on the coordinate subspace $\mathbf{R}^n$ and $v_{1,i}$ is the projection of $\mathbf{v}_i$ on the orthogonal complement to this subspace in $\mathbf{R}^{n+1}$ space [7]. Finally, λ_i is the corresponding eigenvalue. In this way, at any iteration, say k, the RFO method can be written as:

1. Given $\mathbf{g}_k$, $\mathbf{H}_k$ and $\mathbf{S}_k$, solve the corresponding eigenvalue Eq. (4).
2. Select an eigenpair of the set $\{\lambda_i^k, \mathbf{v}_i^k\}_{i=1}^{n+1}$.
3. If $v_{1,i}^k$ differs from zero, then the vector $(1\ \Delta\mathbf{q}_k^{\mathrm{T}})$ is obtained in the following way

$$\begin{pmatrix} 1 \\ \Delta\mathbf{q}_k \end{pmatrix} = \frac{1}{v_{1,i}^k} \begin{pmatrix} v_{1,i}^k \\ \mathbf{v}_{n,i}^k \end{pmatrix} \tag{5}$$

and set $\mathbf{q}_{k+1} = \mathbf{q}_k + \Delta\mathbf{q}_k$.

4. Estimate the quadratic change of the energy due to the displacement $\Delta\mathbf{q}_k$

$$\mathrm{Q}(\Delta\mathbf{q}_k) = \frac{1}{2}\frac{\lambda_i^k}{\left(v_{1,i}^k\right)^2} = \frac{1}{2}\lambda_i^k(1 + \Delta\mathbf{q}_k^{\mathrm{T}}\mathbf{S}_k\Delta\mathbf{q}_k) \tag{6}$$

In the last equation, the normalization condition

$$\left(v_{1,i}^k\right)^2 + \left(\mathbf{v}_{n,i}^k\right)^{\mathrm{T}}\mathbf{S}_k\mathbf{v}_{n,i}^k = 1 \tag{7}$$

has been employed. In step 1, one normally takes $\mathbf{S}_k = \mathbf{I}$ for each iteration. Step 2 needs some comments. At any iteration, if one is interested in locating a minimum, the eigenpair to be selected is $i = 1$. If we want to locate a maximum, $i = n + 1$, while for a saddle point of order p, $i = p + 1$ [2, 7, 8, 10]. We are mainly interested in saddle points of order $p = 1$ which are associated with true transition states; in this case, $i = 2$. Finally, regarding Eq. (5), we see that the displacement vector $\Delta\mathbf{q}_k$ is a function of the selected eigenvector.

2.2 Analysis of the RFO method and selection of the $\mathbf{S}$ *matrix*

This analysis is based on the results of Banerjee et al. [2]. Multiplying Eq. (4) by the inverse of $v_{1,i}$ and taking into account Eq. (5) after arrangement we get

$$\mathbf{g}^{\mathrm{T}}\Delta\mathbf{q} = \lambda_i \tag{8a}$$

$$\mathbf{g} + (\mathbf{H} - \lambda_i\mathbf{S})\Delta\mathbf{q} = \mathbf{0} \tag{8b}$$

Expressions (8) are the partitioned equations of the RFO method [2]. One can see Eq. (8b) as a shifted Newton-Raphson formula. Using Eq. (8b) one computes $\Delta\mathbf{q}$; using Eq. (8a) one evaluates the shift λ_i. In other words, the RFO method is a Newton-Raphson method with a level shift where the calculation of the shift parameter is well defined [10]. However, there still exists an undetermination on the selection of the $\mathbf{S}$ matrix. This matrix can be seen as a correction of the Hessian matrix $\mathbf{H}$ in the region where the quadratic approximation, $\mathrm{Q}(\Delta\mathbf{q})$,given by Eq. (1), fails. The question now is how to select the $\mathbf{S}$ matrix. With the aim of improving the Newton-Raphson method, Fletcher [6] introduced the restricted step technique, whereby the step $\Delta\mathbf{q}$ is restricted within the region where $\mathrm{Q}(\Delta\mathbf{q})$ agrees with the $E(\mathbf{q} + \Delta\mathbf{q}) - E(\mathbf{q})$ function. This region is the trust region characterized by a trust radius, R. In a normal Newton-Raphson method the application of the restricted step technique results in a formula which is similar to Eq. (8b), but taking $\mathbf{S} = \mathbf{I}$ and the λ_i parameter as a Lagrange multiplier, this ensures [6]

$$\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q} = R^2 \tag{8c}$$

and, consequently, is evaluated in a very different way rather than using Eq. (8a). In the RFO method the parameter λ_i does not ensure that $\Delta\mathbf{q}$ falls into the trust region because it is computed through Eq. (8a). The only way that the RFO step solution, $\Delta\mathbf{q}$, falls into the trust

region is by an adequate selection of the **S** matrix such that satisfies eq. (8c). However, the matrix **S** possesses $n(n+1)/2$ unknowns because it is symmetric, so we have more unknown parameters that equations. This problem can be solved by taking

$$\mathbf{S} = \alpha \mathbf{I} \tag{8d}$$

in the representation in which the Hessian matrix **H** is diagonal and $\alpha > 0$. The set of Eqs. (8) are the basis of the restricted step RFO (RS-RFO) algorithm.

The analysis of the RS-RFO method will be carried out in the representation of the set of eigenvectors of the Hessian matrix **H**, $\{\mathbf{w}_j\}_{j=1}^n$. In this basis $\Delta\mathbf{q}$ is

$$\Delta\mathbf{q} = -\sum_{i=1}^{n} \frac{(\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}}{h_i - \lambda\alpha}\mathbf{w}_i = -\sum_{i=1}^{n} \frac{(\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}}{h_i - v}\mathbf{w}_i \tag{9}$$

where $v = \lambda\alpha$. Consequently $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ is

$$\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q} = \sum_{i=1}^{n} \left(\frac{(\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}}{h_i - v} \right)^2 \tag{10}$$

where $\{h_j\}_{j=1}^n$ are the set of eigenvalues of the **H** matrix. Using the same representation, Eqs. (8a) and (8b) can be rearranged as

$$\lambda = -\sum_{i=1}^{n} \frac{\left((\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}\right)^2}{h_i - v} = F(v) = \frac{v}{\alpha} \tag{11}$$

Note that for $\alpha = 1$ it follows that $v = \lambda$. The parametric representation with respect to v of Eqs. (10) and (11) is given in Fig. 1. The axes $F(v)$ and v go from $-\infty$ to $+\infty$ and the axis $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ goes form 0 to $+\infty$. The function $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ has asymptotes at the eigenvalues of the Hessian matrix **H**. In any minimization process, which corresponds to select the lowest eigenpair of eigenvalue Eq. (4), the function $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ goes from 0 to $+\infty$ as v goes from $-\infty$ to h_1. This means that there exists a v and consequently a λ and α such that Eq. (8c) is satisfied. In Fig. 1, v_1^1 is the solution of the RS-RFO equations for a minimization because for this v value the function $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ intersects with the straight line $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q} = R^2$. Note that this v_1^1 value is obtained by a $\alpha = \alpha_1$. In searching a transition state, the RFO method converges if v is taken to be within the interval $h_1 < v < h_2$. However, there is no guaranty that the function $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ in this interval possesses a common point with the straight line $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q} = R^2$. In Fig. 1, no value of v such that $h_1 < v < h_2$ can satisfy $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q} = R^2$, since in this interval no point of the function $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ intersects with this straight line. These results show an important limitation of the RS-RFO method to locate transition states. These drawbacks have been pointed out many times [2, 5].

From the last discussion, one can see that for a given vector **g** and matrix **H**, the RS-RFO equations depend uniquely on the independent parameter α. For a value α, through Eq. (4) we obtain one and only one set of $n+1$ eigenpairs; selecting one eigenvector from this set and using Eq. (5) gives a displacement vector $\Delta\mathbf{q}$ that may or

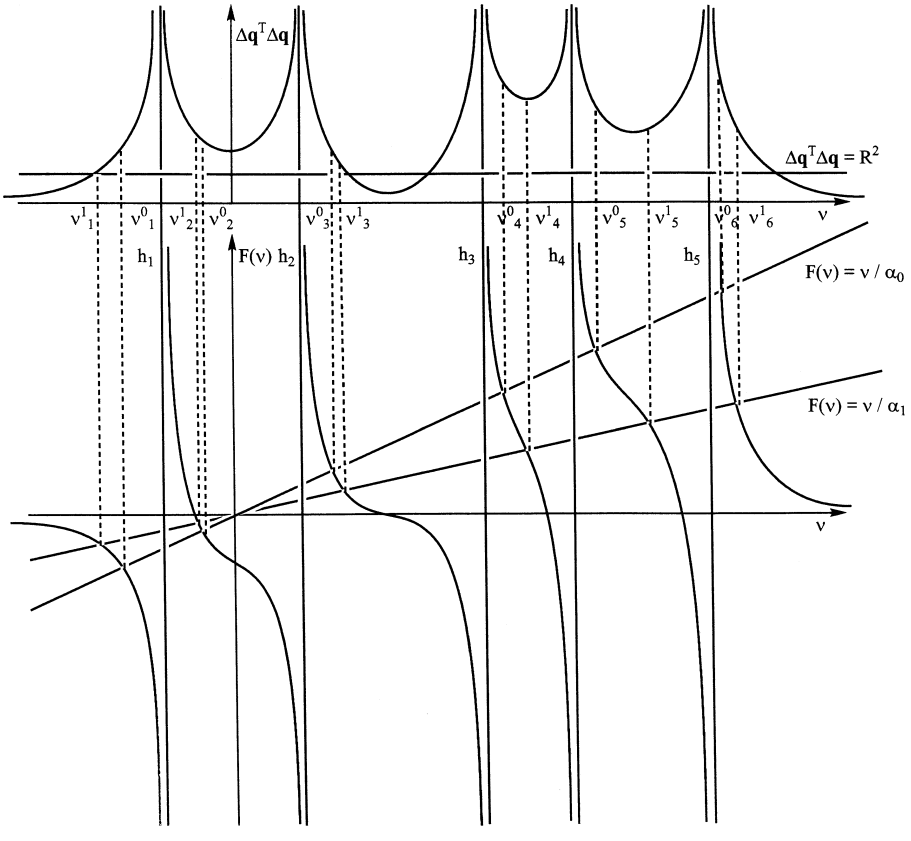

Fig. 1. Schematic plots of $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ and $F(v)$ with respect to v given by Eqs. (10) and (11), respectively. Both plots are coupled in order to see the behavior of the restricted-step RFO method. The set $h_1, h_2, \ldots$ is the set of eigenvalues of the Hessian matrix **H**. The sets, $v_1^0, v_2^0, \ldots$ and $v_1^1, v_1^1, \ldots$ are the shift parameters of the Hessian matrix. Each set of the shift parameter is associated with a different α value. Note that $v_i^j = \lambda_i \alpha_j$, where the set $\lambda_1, \lambda_2, \ldots$ are the eigenvalues of the augmented Hessian matrix for a given α. The parameter R is the trust radius. In the normal RFO method, we take the shift parameter v_1 for minima searches and v_2 for the location of transition states

may not satisfy Eq. (8c). Owing to this fact, it is important to see the analytical dependency of $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ with respect to the parameter α in order to write an efficient RS-RFO algorithm. First we differentiate $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ with respect to the paramete α, that is

$$\frac{d(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})}{d\alpha} = 2\frac{d(\lambda\alpha)}{d\alpha}\sum_{i=1}^{n}\frac{\left((\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}\right)^2}{(h_i-\lambda\alpha)^3} \tag{12}$$

where Eq. (10) has been used. Now we need the derivative of λ with respect to α. Differentiating with respect to α the first two terms of Eq. (11) we get

$$\begin{aligned}\frac{d\lambda}{d\alpha} &= -\frac{d(\lambda\alpha)}{d\alpha}\sum_{i=1}^{n}\frac{\left((\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}\right)^2}{(h_i-\lambda\alpha)^2}\\ &= -\frac{d(\lambda\alpha)}{d\alpha}\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}\\ &= -\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}\left(\frac{d\lambda}{d\alpha}\alpha+\lambda\right)\end{aligned} \tag{13}$$

where Eq. (10) has been used and the fact that $v = \lambda\alpha$. From the last equation we obtain

$$\frac{d\lambda}{d\alpha} = -\frac{\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}}{(1+\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}\alpha)}\lambda \tag{14}$$

Now we can evaluate $d(\lambda\alpha)/d\alpha$:

$$\frac{d(\lambda\alpha)}{d\alpha} = \frac{d\lambda}{d\alpha}\alpha+\lambda = \frac{\lambda}{(1+\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}\alpha)} \tag{15}$$

where Eq. (14) has been used. Substituting Eq. (15) into Eq. (12) we get

$$\frac{d(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})}{d\alpha} = 2\frac{\lambda}{(1+\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}\alpha)}\sum_{i=1}^{n}\frac{\left((\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}\right)^2}{(h_i-\lambda\alpha)^3} \tag{16}$$

Equation (16) is the main equation in order to write an algorithm based on the RS-RFO equations and it gives the analytical dependency of $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ with respect to the α parameter.

2.3 The restricted step partitioned RFO method to locate saddle points

As pointed out before, the RFO method presents some difficulties to locate saddle points. Owing to this fact, Banerjee et al. [2] proposed a slight modification of the method. Based on the idea of Bell et al. [11], they partitioned the quadratic approximation model given in Eq. (1) using the conjugate directions, specifically the eigenvectors of the Hessian matrix. In this way, one has two quadratic approximations. One is the quadratic approximation defined using the eigenvectors set of the **H** matrix along the directions the energy is maximized, and the other taking the set of the eigenvectors of the **H** matrix along the directions the energy is minimized. If one applies independently the RFO method to each one of these two quadratic approximations, one has the so-called partitioned RFO (P-RFO) method [2]. Using this method, one has two eigenvalues and two displacement vectors, one associated with the maximization subspace and the other associated with the minimization subspace. Note that the eigenvalue of the maximization subspace is the highest eigenvalue of the RFO eigenvalue equation of this subspace.

Let us analyze the restricted step method applied to the P-RFO algorithm to locate a saddle point of order p (RS-P-RFO). In this case, Eq. (10) takes the following form:

$$\begin{aligned}\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q} &= (\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\max} + (\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\min}\\ &= \sum_{i=1}^{p}\left(\frac{(\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}}{h_i-v_{\max}}\right)^2 + \sum_{i=p+1}^{n}\left(\frac{(\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}}{h_i-v_{\min}}\right)^2\\ &= \sum_{i=1}^{p}\left(\frac{(\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}}{h_i-\lambda_{\max}\alpha}\right)^2 + \sum_{i=p+1}^{n}\left(\frac{(\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}}{h_i-\lambda_{\min}\alpha}\right)^2\end{aligned} \tag{17}$$

where the subindexes max and min denote the maximization and minimization subspaces, respectively. As the parameter α increases (decreases), $v_{\max}$ increases (decreases) and consequently $(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\max}$ decreases (increases). In the same way, as α increases (decreases), $v_{\min}$ decreases (increases) and, consequently, $(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\min}$ decreases (increases). This fact can be observed by analyzing Eq. (16) for this situation:

$$\begin{aligned}\frac{d(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})}{d\alpha} &= \frac{d(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\max}}{d\alpha} + \frac{d(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\min}}{d\alpha}\\ &= 2\frac{\lambda_{\max}}{(1+(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\max}\alpha)}\sum_{i=1}^{p}\frac{\left((\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}\right)^2}{(h_i-\lambda_{\max}\alpha)^3}\\ &\quad + 2\frac{\lambda_{\min}}{(1+(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\min}\alpha)}\sum_{i=p+1}^{n}\frac{\left((\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}\right)^2}{(h_i-\lambda_{\min}\alpha)^3}\\ &= 2\frac{v_{\max}}{(1+(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\max}\alpha)\alpha}\sum_{i=1}^{p}\frac{\left((\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}\right)^2}{(h_i-v_{\max})^3}\\ &\quad + 2\frac{v_{\min}}{(1+(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\min}\alpha)\alpha}\sum_{i=p+1}^{n}\frac{\left((\mathbf{w}_i)^{\mathrm{T}}\mathbf{g}\right)^2}{(h_i-v_{\min})^3}\end{aligned} \tag{18}$$

where $d(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\max}/d\alpha \le 0$ since $v_{\max} > h_p$ and h_p is the highest eigenvalue of the max subspace and $v_{\max} > 0$. On the other hand, $d(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})_{\min}/d\alpha \le 0$ since $v_{\min} < h_{p+1}$ and h_{p+1} is the lowest eigenvalue of the min subspace and $v_{\min} < 0$. From the above analysis we conclude that the overall $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ goes from $+\infty$ to 0 as α changes from 0 to $+\infty$. This means that there exists a $\lambda_{\max}$ and $\lambda_{\min}$ for a given α such that Eqs. (8) will be satisfied for the partitioned RFO, which is the basis of the RS-P-RFO method.

Finally, we say that the RS-P-RFO algorithm forces a decoupling between the variables of the max and min subspaces. The importance of this decoupling in the

optimization process and the reformulation of a RS-RFO algorithm to locate transition states will be analyzed below.

3 Computational implementation, examples, and discussion

3.1 Details of the algorithm

In this subsection we give the organization of the steps involved in the iterative procedure of the proposed RS-RFO/RS-P-RFO algorithm. The description of the algorithm is mainly focused to locate saddle points. However, its extension to minima and maxima is trivial.

3.1.1 Initialization

Given $\mathbf{q}_0$, a trust radius R_0, calculate $E_0 = E(\mathbf{q}_0)$, the gradient $\mathbf{g}_0$, and the Hessian matrix $\mathbf{H}_0$. Set $k = 0$.

3.1.2 Analysis of the Hessian

Compute the eigenpairs of the Hessian matrix $\mathbf{H}_k, \{h_i^k, \mathbf{w}_i^k\}_{i=1}^n$, project the gradient vector on the eigenvector basis, $\{(\mathbf{w}_i^k)^{\mathrm{T}}\mathbf{g}_k = f_i^k\}_{i=1}^n$. Select the transition vector, that is, the eigenvector along which the energy is maximized: $\mathbf{w}_{\mathrm{tv}}^k, \mathrm{tv} \in [1, n]$. Except in the first iteration, the selection of the transition vector is done following the idea of Simons et al. [4].

3.1.3 Solution of the RS-P-RFO equations

This step involves a micro-iterative process.

1. Select $\alpha_0 = 1$, set $\mu = 0$.
2. As for the normal P-RFO algorithm [2], build and solve the next two eigenvalue equations:

$$\begin{pmatrix} 0 & f_{\mathrm{tv}}^k \\ f_{\mathrm{tv}}^k & h_{\mathrm{tv}}^k \end{pmatrix} \begin{pmatrix} {}_{\max}v_1^k \\ {}_{\max}v_2^k \end{pmatrix} = \lambda_{\max}^k \begin{pmatrix} 1 & 0 \\ 0 & \alpha_\mu \end{pmatrix} \begin{pmatrix} {}_{\max}v_1^k \\ {}_{\max}v_2^k \end{pmatrix} \tag{19a}$$

$$\begin{pmatrix} 0 & (\mathbf{f}_{n-1}^k)^{\mathrm{T}} \\ \mathbf{f}_{n-1}^k & \mathbf{H}_{\mathrm{D}}^k \end{pmatrix} \begin{pmatrix} {}_{\min}v_1^k \\ {}_{\min}\mathbf{v}_{n-1}^k \end{pmatrix} = \lambda_{\min}^k \begin{pmatrix} 1 & 0 \\ 0 & \alpha_\mu \end{pmatrix} \begin{pmatrix} {}_{\min}v_1^k \\ {}_{\min}\mathbf{v}_{n-1}^k \end{pmatrix} \tag{19b}$$

where $\mathbf{f}_{n-1}^k$ is the vector $\mathbf{f}^k$ without the element f_{tv}^k and the matrix $\mathbf{H}_{\mathrm{D}}^k$ is a diagonal matrix formed by the eigenvalues of the matrix $\mathbf{H}_k$ except h_{tv}^k.

3. From Eq. (19a), take the eigenpair $(\lambda_{\max}^k, {}_{\max}\mathbf{v}^k)$ of highest eigenvalue, and from Eq. (19b) take the eigenpair $(\lambda_{\min}^k, {}_{\min}\mathbf{v}^k)$ of the lowest eigenvalue. Using these two eigenvectors and Eq. (5), compute Δq_{tv}^k and $\Delta\mathbf{q}_{n-1}^k$, where $\Delta\mathbf{q}_{n-1}^k$ is the displacement vector $\Delta\mathbf{q}^k$ without the element Δq_{tv}^k.
4. If $\mu = 0$ and $(\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k)^{1/2} > R_k$ or $\mu \neq 0$ and $|(\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k)^{1/2} - R_k| > \varepsilon_1$ and $|\alpha_{\mu-1} - \alpha_\mu| > \varepsilon_2$, then compute the new $\alpha_{\mu+1}$ using the expression

$$\alpha_{\mu+1} = \alpha_\mu + \frac{2\left(R_k(\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k)^{1/2} - \Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k\right)}{\left(\dfrac{d\left(\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k\right)}{d\alpha}\right)_{\alpha=\alpha_\mu}} \tag{20}$$

where the derivative $d(\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k)/d\alpha$ is evaluated according to Eq. (18); set $\mu = \mu + 1$ and go to step 2; otherwise end the micro-iterative process, since the vector $\Delta\mathbf{q}_k$ satisfies the RS-P-RFO equations.

3.1.4 Prediction of the energy variation

Using Eq. (6), estimate the quadratic change in energy:

$$\begin{aligned} \mathrm{Q}(\Delta\mathbf{q}_k) &= \mathrm{Q}\left(\Delta q_{\mathrm{tv}}^k\right) + \mathrm{Q}\left(\Delta\mathbf{q}_{n-1}^k\right) \\ &= \frac{1}{2}\left[\frac{\lambda_{\max}^k}{({}_{\max}v_1^k)^2} + \frac{\lambda_{\min}^k}{({}_{\min}v_1^k)^2}\right] \end{aligned} \tag{21}$$

3.1.5 Evaluation of the trust region

Compute the energy at the point $\mathbf{q}_{k+1} = \mathbf{q}_k + \Delta\mathbf{q}_k, E_{k+1} = E(\mathbf{q}_{k+1})$, and the coefficient $r = (E_{k+1} - E_k)/\mathrm{Q}(\Delta\mathbf{q}_k)$. If $r < r_{\mathrm{l}}^{\mathrm{e}}$ or $r > r_{\mathrm{u}}^{\mathrm{e}}$, set $R_{k+1} = (\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k)^{1/2}/\mathrm{Sf}$. If $r_{\mathrm{l}}^{\mathrm{i}} \leq r \leq r_{\mathrm{u}}^{\mathrm{i}}$ and $|(\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k)^{1/2} - R_k| \leq \varepsilon_1$, set $R_{k+1} = R_k\,(\mathrm{Sf})^{1/2}$.

3.1.6 Test on the displacement vector $\Delta\mathbf{q}_k$ and convergence criteria

If $r > \mathrm{Ub}$ or $r < \mathrm{Lb}$, then compute a new $\Delta\mathbf{q}_k$ at the same point $\mathbf{q}_k$ but using the new R_{k+1} trust radius and set $\mathbf{q}_{k+1} = \mathbf{q}_k, E_{k+1} = E_k, k = k + 1$, and go to Sect. 3.1.3, otherwise check the convergence criteria: $(\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k/n)^{1/2} \leq \varepsilon_3$, $(\Delta\mathbf{g}_k^{\mathrm{T}}\Delta\mathbf{g}_k/n)^{1/2} \leq \varepsilon_4$, $|(\Delta\mathbf{q}_k)_i| \leq \varepsilon_5$, $|(\Delta\mathbf{g}_k)_i| \leq \varepsilon_6, i = 1, n$. If they are satisfied, then stop. The point $\mathbf{q}_k$ is a stationary point.

If the convergence criteria are not satisfied, then compute $\mathbf{q}_{k+1} = \mathbf{q}_k + \Delta\mathbf{q}_k$, and the gradient vector at this point, $\mathbf{g}_{k+1}$. Update the approximate Hessian matrix, $\mathbf{H}_{k+1}$, using the Murtagh-Sargent-Powell (MSP) formula [9c, 12]. Set $k = k + 1$ and go to Sect. 3.1.2.

3.2 Algorithm comments

The above algorithm is quite close to that proposed for the QA/TRIM and RQNR [9] methods. The parameters $r_{\mathrm{l}}^{\mathrm{e}}, r_{\mathrm{l}}^{\mathrm{i}}, r_{\mathrm{u}}^{\mathrm{e}}$ and $r_{\mathrm{u}}^{\mathrm{i}}$ lie in the range $\mathrm{Lb} \leq r_{\mathrm{l}}^{\mathrm{e}} < r_{\mathrm{l}}^{\mathrm{i}} < r_{\mathrm{u}}^{\mathrm{i}} < r_{\mathrm{u}}^{\mathrm{e}} \leq \mathrm{Ub}$, where $r_{\mathrm{l}}^{\mathrm{e}} = \mathrm{Lb} + \delta_e$, $r_{\mathrm{l}}^{\mathrm{i}} = \mathrm{Lb} + \delta_{\mathrm{i}}$, $r_{\mathrm{u}}^{\mathrm{e}} = \mathrm{Ub} - \delta_{\mathrm{e}}$, $r_{\mathrm{u}}^{\mathrm{i}} = \mathrm{Ub} - \delta_{\mathrm{i}}$ with $\mathrm{Ub} \geq \delta_{\mathrm{i}} > \delta_{\mathrm{e}} \geq \mathrm{Lb}$ and $\mathrm{Ub} > 1 > \mathrm{Lb}$. Suggested values of these parameters are: $\mathrm{Lb} = 0, \mathrm{Ub} = 2, \delta_{\mathrm{e}} = 0.75$, and $\delta_{\mathrm{i}} = 0.80$. The parameter Sf is a scaling factor with a suggested value of 2. The initial trust radius is $R_0 = 0.03$ Å/rad.

The diagonalization of the generalized eigenvalue Eqs. (19) can be carried out using standard methods by transforming them to the simplest eigenvalue equation:

$$\begin{pmatrix} 0 & \alpha_\mu^{-1/2}(\mathbf{f}_{n-1}^k)^{\mathrm{T}} \\ \alpha_\mu^{-1/2}\mathbf{f}_{n-1}^k & \alpha_\mu^{-1}\mathbf{H}_{\mathrm{D}}^k \end{pmatrix} \begin{pmatrix} \min z_1^k \\ \min \mathbf{z}_{n-1}^k \end{pmatrix} = \lambda_{\min}^k \begin{pmatrix} \min z_1^k \\ \min \mathbf{z}_{n-1}^k \end{pmatrix} \tag{22}$$

and by back transformation using the equation

$$\begin{pmatrix} \min v_1^k \\ \min \mathbf{v}_{n-1}^k \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \alpha_\mu^{-1/2} \end{pmatrix} \begin{pmatrix} \min z_1^k \\ \min \mathbf{z}_{n-1}^k \end{pmatrix} \tag{23}$$

Equation (20) used in step 4 of Sect. 3.1.3 needs some comments. Since we are interested in the solution of the nonlinear Eq. (8c), where $\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k$ is a function of α, this is carried out using the Newton method to find the zero value of the function $\Phi(\alpha) = 1 - (\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k)^{1/2}/R_k$, which in this case takes the following expression:

$$\begin{aligned} \alpha_{\mu+1} &= \alpha_\mu - \frac{\Phi(\alpha)_{\alpha=\alpha_\mu}}{\left(\dfrac{d\Phi(\alpha)}{d\alpha}\right)_{\alpha=\alpha_\mu}} \\ &= \alpha_\mu + \frac{2\left(R_k(\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k)^{1/2} - \Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k\right)}{\left(\dfrac{d\left(\Delta\mathbf{q}_k^{\mathrm{T}}\Delta\mathbf{q}_k\right)}{d\alpha}\right)_{\alpha=\alpha_\mu}} \end{aligned} \tag{24}$$

and justifies Eq. (20). This procedure can be seen as an extension of the Hebden technique [6, 13].

3.3 Numerical results and performance of the algorithm

The numerical analysis and behavior of the algorithm presented above (RS-P-RFO) is carried out for the location and optimization of the transition structures of several reactions. These reactions were studied using the appropriate wave function and the AM1 semiempirical Hamiltonian [14] implemented in a modified version of the MOPAC program [15]. The convergence criteria are $\varepsilon_1 = 10^{-4}$ Å/rad, $\varepsilon_2 = 10^{-5}$ (Å)$^{-2}$/(rad)$^{-2}$, $\varepsilon_3 = 6 \times 10^{-4}$ Å/rad, $\varepsilon_4 = 4 \times 10^{-2}$ kcal (mol Å)$^{-1}$/kcal (mol rad)$^{-1}$, $\varepsilon_5 = 10^{-3}$ Å/rad, and $\varepsilon_6 = 5 \times 10^{-2}$ kcal (mol Å)$^{-1}$/kcal (mol rad)$^{-1}$. The parameters ε_1 and ε_2 are the convergence criteria used in the algorithm to solve the RS-P-RFO equations. On the other hand, the parameters $\varepsilon_3, \varepsilon_4, \varepsilon_5$, and ε_6 are the normal convergence criteria used in any optimization algorithm. Specifically, ε_3 and ε_5 are the convergence criteria on the norm and the biggest component in the absolute value of the $\Delta\mathbf{q}_k$ vector, whereas ε_4 and ε_6 are the corresponding criteria for the $\Delta\mathbf{g}_k$ vector. The Z-matrix geometrical coordinates were used during the process. However, according to the proposed algorithm, the optimization of the quadratic model is carried out using the coordinates in that the Hessian matrix is diagonal. The initial Z-matrices were taken from [16–19]. The results are shown in Table 1. The behavior of the present algorithm is also compared with a P-RFO method which differs from the proposed RS-P-RFO technique in that the resulting displacement vector, $\Delta\mathbf{q}$, is scaled by some factor if $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ is larger than the current trust radius. In general, RS-P-RFO converges much better than the P-RFO method. Note that in examples 12 and 16 in Table 1 the P-RFO technique does not converge to the correct stationary point but to a minimum. This never occurs when the RS-P-RFO algorithm is used. In examples 8 and 9 the RS-P-RFO technique does not converge within 100 iterations and the same behavior is observed in example 9 using the P- RFO algorithm. As we will see below, this is due to the partition of the AH matrix. We conclude from these numerical examples that, in general, it is much better to take an algorithm such that the displacement vector, $\Delta\mathbf{q}$, corresponds to the optimum solution of the RFO equations within the trust radius, rather than just scaling it by some factor.

Finally, in Table 2 we show the performance of the micro-iterative procedure described in Sect. 3.1.3 of the proposed algorithm for some reactions presented in Table 1. We observe that when the molecular geometry is far from the optimum solution, if the trust radius is quite big, say ≈ 0.17 Å/rad, the number of micro-iterations needed to converge is around 7. However, in the same situation but with a small trust radius, say ≈ 0.05 Å/rad, the number of iterations needed to reach the convergence of the micro-iterative process is around 13. Near to the converged molecular geometry, since the length of the displacement vector, that is $(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})^{1/2}$, is much smaller that the current trust radius, the micro-iterative process is off. In some situations and normally when the trust radius is small, the derivative $d\Phi(\alpha)/d\alpha$ is very small, so Eq. (24) cannot be used. In these cases we use the secant method [20], which consists of substituting in Eq. (24) the derivative by the quotient $(\Phi(\alpha_\mu) - \Phi(\alpha_{\mu-1}))/(\alpha_\mu - \alpha_{\mu-1})$. Looking at Fig. 1, it is easy to understand why in these cases the derivative $d\Phi(\alpha)/d\alpha$ is small. For a small trust radius, the function $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}$ with respect to v associated with the first eigenpair intersects with the function $\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q} = R^2$ when it presents an asymptotic behavior; this implies that the function $\Phi(\alpha)$ is almost constant in this region and, consequently, $d\Phi(\alpha)/d\alpha \approx 0$.

4 The relation between RS-RFO and the QA/TRIM and RQNR algorithms to locate saddle points; a unified treatment. The RS-I-RFO method

According to the previous discussion, using the RFO methodology to locate a saddle point one needs to partition the variables space into two subspaces. This partition implies that during the quadratic optimization the interaction between both subspaces is not taken into account. The following question arises: is it possible to reformulate the RS-RFO method in such a way that without any type of partition one can find transition structures efficiently? To answer this question we use the concept of the image function of a quadratic potential surface introduced by Smith [21], implemented using restricted step algorithms [9], and analyzed by Sun and Ruedenberg [22].

Let us take the quadratic function, $Q(\Delta\mathbf{q})$, given in Eq. (1). We assume that the Hessian matrix $\mathbf{H}$ possesses

Table 1. Comparison of the number of steps required to optimize transition state geometries

Reaction	Method		
	P-RFO[a]	RS-P-RFO[b]	RS-I-RFO[c]
1. $CH_3O \rightarrow CH_2OH$ [9c, 17]	14	18	19
2. Cyclopropyl radical ring opening [9c]	15	16	15
3. Bicyclo[2.2.1]butane ring opening, TS 1 [9c]	15	17	16
4. Bicyclo[2.2.1]butane ring opening, TS 2 [9c]	28	24	26
5. β-(formyloxy)ethyl radical ring opening [9c]	25	13	13
6. $HCONHOH \rightarrow HCOHNHO$ [18]	12	10	10
7. Claisen rearrangement of $CH_2{=}CHCH_2OCH{=}CH_2$ [16, 18]	42	36	33
8. Retro-Diels-Alder reaction of cyclohexene [18]	24	Fails[d]	21
9. $CH_3O_2H \rightarrow CH_3 + O_2H$ [17, 19]	Fails[d]	Fails[d]	37
10. Cyclopropylcarbene → Cyclobutene [17, 19]	47	35	27
11. Tetrazine decomposition [18, 19]	31	32	34
12. $CH_3CH_3 \rightarrow CH_2CH_2 + H_2$ [18]	Fails[e]	15	14
13. $OCHNH_3^+ \rightarrow NH_4^+ + CO$ [18]	14	14	18
14. $H_2CNH \rightarrow HCNH_2$ [18]	22	20	23
15. $HCN + H_2 \rightarrow H_2CNH$ [18]	25	22	22
16. $HCNH_2 \rightarrow HCN + H_2$ [18]	Fails[e]	56	53

All optimizations started at the *Z*-matrices given in the respective references; an exact Hessian matrix was calculated at the first iteration only

[a] The P-RFO solution is scaled according the current trust radius

[b] The P-RFO solution is optimized within the trust region; see text for more details

[c] The quadratic model optimized corresponds to the image function of the quadratic model of the objective function. The I-RFO solution is optimized within the trust region; see text for more details

[d] The process did not converge within 100 iterations

[e] The process converged to a minimum

Table 2. Number of micro-iterations needed for each RS-P-RFO iteration of some reactions presented in Table 1

Iteration[b]	Reaction[a]			
	1	5	6	12
1	12 (0.030)	10 (0.030)	13 (0.030)	15 (0.030)
2	15 (0.011)	9 (0.042)	11 (0.042)	15 (0.042)
3	17 (0.015)	8 (0.060)	10 (0.060)	14 (0.060)
4	14 (0.021)	6 (0.085)	9 (0.085)	13 (0.085)
5	14 (0.030)	0 (0.120)	7 (0.120)	7 (0.120)
6	13 (0.042)	0 (0.120)	0 (0.170)	6 (0.170)
7	12 (0.060)	19 (0.008)	0 (0.170)	0 (0.240)
8	12 (0.085)	20 (0.006)	0 (0.006)	4 (0.031)
9	11 (0.120)	21 (0.006)	0 (0.006)	18 (0.008)
10	9 (0.170)	20 (0.006)	0 (0.006)	7 (0.011)
11	4 (0.170)	4 (0.008)		0 (0.006)
12	0 (0.170)	0 (0.012)		0 (0.006)
13	0 (0.034)	0 (0.012)		0 (0.006)
14	0 (0.034)			0 (0.006)
15	0 (0.006)			0 (0.006)
16	0 (0.006)			
17	0 (0.006)			
18	0 (0.006)			

[a] Reaction numbering as in Table 1. The current trust radius in Å rad are given in parentheses

[b] Iteration RS-P-RFO

one and only one negative eigenvalue, that is for $h_{\mathrm{tv}} < 0$ and $h_i > 0$ for $i \neq \mathrm{tv}, i = 1, n$. Following Sun and Ruedenberg [22] we define an image function of $Q(\Delta\mathbf{q})$, denoted by $Q^*(\Delta\mathbf{q})$, by the equation

$$\mathbf{g}^* = \left(\mathbf{I} - 2\mathbf{w}_{\mathrm{tv}}\mathbf{w}_{\mathrm{tv}}^{\mathrm{T}}\right)\mathbf{g} \tag{25}$$

Which is an elementary Householder orthogonal transformation applied on the $\mathbf{g}$ vector. Differentiating Eq. (25) with respect to the geometrical variables we get

$$\begin{aligned}\mathbf{H}^* &= (\mathbf{I} - 2\mathbf{w}_{\mathrm{tv}}\mathbf{w}_{\mathrm{tv}}^{\mathrm{T}})\mathbf{H} = \mathbf{H}(\mathbf{I} - 2\mathbf{w}_{\mathrm{tv}}\mathbf{w}_{\mathrm{tv}}^{\mathrm{T}}) \\ &= h_{\mathrm{tv}}^*\mathbf{w}_{\mathrm{tv}}\mathbf{w}_{\mathrm{tv}}^{\mathrm{T}} + \sum_{\substack{i=1 \\ i\neq \mathrm{tv}}}^{n} h_i\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}}\end{aligned} \tag{26}$$

where $h_{\mathrm{tv}}^* = -h_{\mathrm{tv}}$. From Eq. (25) we see that $f_{\mathrm{tv}}^* = \mathbf{w}_{\mathrm{ts}}^{\mathrm{T}}\mathbf{g}^* = -\mathbf{w}_{\mathrm{ts}}^{\mathrm{T}}\mathbf{g} = -f_{\mathrm{tv}}$ and $f_i^* = \mathbf{w}_i^{\mathrm{T}}\mathbf{g}^* = \mathbf{w}_i^{\mathrm{T}}\mathbf{g} = f_i$, for $i \neq \mathrm{tv}$, $i = 1, n$. It is easy to prove that the stationary point of quadratic function $Q(\Delta\mathbf{q})$ is the same point of the image

quadratic function $Q^*(\Delta\mathbf{q})$ [22]. With this important result in mind one can apply the RS-RFO method to find the minimum of the image quadratic function $Q^*(\Delta\mathbf{q})$, since $\mathbf{H}^*$ is positive definite, to locate a saddle point. Taking the eigenpair corresponding to the lowest eigenvalue, λ^*, of the RFO eigenvalue equation of the image quadratic function, the vector displacement $\Delta\mathbf{q}$ is

$$\begin{aligned}\Delta\mathbf{q} = \Delta\mathbf{q}^* &= -\frac{f_{\mathrm{tv}}^*}{h_{\mathrm{tv}}^* - \lambda^*\alpha}\mathbf{w}_{\mathrm{ts}} - \sum_{\substack{i=1\\ i\neq \mathrm{tv}}}^{n}\frac{f_i}{h_i - \lambda^*\alpha}\mathbf{w}_i \\ &= -\frac{f_{\mathrm{tv}}}{h_{\mathrm{tv}} + \lambda^*\alpha}\mathbf{w}_{\mathrm{ts}} - \sum_{\substack{i=1\\ i\neq \mathrm{tv}}}^{n}\frac{f_i}{h_i - \lambda^*\alpha}\mathbf{w}_i \\ &= -\frac{f_{\mathrm{tv}}}{h_{\mathrm{tv}} + v^*}\mathbf{w}_{\mathrm{ts}} - \sum_{\substack{i=1\\ i\neq \mathrm{tv}}}^{n}\frac{f_i}{h_i - v^*}\mathbf{w}_i \end{aligned} \tag{27}$$

where $v^* = \lambda^*\alpha$. Equation (27) as function of v^* was given by Helgaker [9a]. The same result is obtained if we use the following transformation:

$$\mathbf{g}^* = \left(\mathbf{I} - 2\sum_{\substack{i=1\\ i\neq \mathrm{tv}}}^{n}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}}\right)\mathbf{g} \tag{28}$$

to define a new image quadratic function. Note that using this transformation $\mathbf{H}^*$ is now negative definite. Applying again the RS-RFO method to find the maximum of this image quadratic function, the resulting vector displacement $\Delta\mathbf{q}$ is

$$\begin{aligned}\Delta\mathbf{q} = \Delta\mathbf{q}^* &= -\frac{f_{\mathrm{tv}}}{h_{\mathrm{tv}} - \lambda^*\alpha}\mathbf{w}_{\mathrm{ts}} - \sum_{\substack{i=1\\ i\neq \mathrm{tv}}}^{n}\frac{f_i^*}{h_i^* - \lambda^*\alpha}\mathbf{w}_i \\ &= -\frac{f_{\mathrm{tv}}}{h_{\mathrm{tv}} - \lambda^*\alpha}\mathbf{w}_{\mathrm{ts}} - \sum_{\substack{i=1\\ i\neq \mathrm{tv}}}^{n}\frac{f_i}{h_i + \lambda^*\alpha}\mathbf{w}_i \\ &= -\frac{f_{\mathrm{tv}}}{h_{\mathrm{tv}} - v^*}\mathbf{w}_{\mathrm{ts}} - \sum_{\substack{i=1\\ i\neq \mathrm{tv}}}^{n}\frac{f_i}{h_i + v^*}\mathbf{w}_i \end{aligned} \tag{29}$$

In this case, λ^* corresponds to the highest eigenvalue of the RFO eigenvalue equation. Since λ^* is the highest eigenvalue, the v^* parameter obeys the next inequality (see Fig. 1), $v^* = \lambda^*\alpha > \max\{h_{\mathrm{tv}}, \{h_i^*\}_{i=1,i\neq \mathrm{tv}}^{n}\}$ and the domain of the parameter v^* in Eq. (29) goes from $+\infty$ to $\max\{h_{\mathrm{tv}}, -\min\{h_i\}_{i=1,i\neq \mathrm{tv}}^{n}\}$. Equation (29) together with this domain of the shift parameter v^* constitute the basis of the QA/TRIM and RQNR algorithms as defined by Culot et al. [9b] and Bofill [9c]. Making the same reasoning, we see that the domain of the shift parameter v^* in Eq. (27) goes from $-\infty$ to $\min\{h_{\mathrm{tv}}^*, \{h_i\}_{i=1,i\neq \mathrm{tv}}^{n}\} = \min\{-h_{\mathrm{tv}}, \min\{h_i\}_{i=1,i\neq \mathrm{tv}}^{n}\}$. These results show that the QA/TRIM and RQNR algorithms [9] and the RS-RFO method applied on an appropriate image quadratic function of the quadratic model of the objective function, hereafter called the RS-I-RFO method, are the same. The difference between them is the way to compute the shift parameter v. For the QA/TRIM and RQNR algorithms [9], the shift parameter v is a Lagrangian multiplier, while in the RS-I-RFO method the parameter v comes from the product of an eigenvalue, λ, associated with a generalized eigenvalue equation with the corresponding metric, α, used in this eigenvalue equation.

The RS-I-RFO algorithm is close to the algorithm described in Sect. 3.1; however, we will comment on their differences. First, in step 2 of Sect. 3.1.3 we have only one eigenvalue equation rather than two. The AH matrix to be diagonalized is built using the $\mathbf{f}^*$ vector and the $\mathbf{H}_{\mathrm{D}}^*$matrix, the diagonal matrix form of $\mathbf{H}^*$, rather than the $\mathbf{f}$ vector and the $\mathbf{H}_{\mathrm{D}}$ matrix, respectively. In step 3 of Sect. 3.1.3, one should take the eigenpair corresponding to the lowest eigenvalue if the image quadratic function is built according to Eq. (25), or the eigenpair corresponding to the highest eigenvalue if the image quadratic function is built using Eq. (28). With the selected eigenvector through Eq. (5) one gets directly the displacement vector $\Delta\mathbf{q}$. In step 4, of Sect. 3.1.3 the derivative that appears in Eq. (20), $d(\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q})/d\alpha$, is evaluated using Eq. (16). Finally, in Sect. 3.1.4, some care should be taken in order to compute $Q(\Delta\mathbf{q})$, since $Q^*(\Delta\mathbf{q}) \neq Q(\Delta\mathbf{q})$. This is due to the fact that λ^*, the selected eigenvalue of the AH matrix associated with the image quadratic function, is different from the corresponding eigenvalue λ of the AH matrix associated with the model quadratic function of the objective function. If the image quadratic function is defined through the Householder transformation $(\mathbf{I} - 2\mathbf{w}_{\mathrm{tv}}\mathbf{w}_{\mathrm{tv}}^{\mathrm{T}})$ [see Eqs. (25) and (26)], the relation between both eigenvalues λ and λ^* is

$$\begin{aligned}2\lambda^* &= \frac{\begin{pmatrix}1 & \Delta\mathbf{q}^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}0 & (\mathbf{g}^*)^{\mathrm{T}}\\ \mathbf{g}^* & \mathbf{H}^*\end{pmatrix}\begin{pmatrix}1\\ \Delta\mathbf{q}\end{pmatrix}}{1 + \alpha\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}} \\ &= \frac{2\mathbf{g}^{\mathrm{T}}\left(\mathbf{I} - 2\mathbf{w}_{\mathrm{tv}}\mathbf{w}_{\mathrm{tv}}^{\mathrm{T}}\right)\Delta\mathbf{q} + \Delta\mathbf{q}^{\mathrm{T}}\mathbf{H}\left(\mathbf{I} - 2\mathbf{w}_{\mathrm{tv}}\mathbf{w}_{\mathrm{tv}}^{\mathrm{T}}\right)\Delta\mathbf{q}}{1 + \alpha\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}} \\ &= 2\lambda - 2\frac{2f_{\mathrm{tv}}\Delta q_{\mathrm{tv}} + h_{\mathrm{tv}}(\Delta q_{\mathrm{tv}})^2}{1 + \alpha\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}} \end{aligned} \tag{30}$$

where Eqs. (2) and (8d) applied to the image quadratic function have been used together with Eqs. (25) and (26). Rearranging Eq. (30), we get

$$\lambda = \lambda^* + \frac{2f_{\mathrm{tv}}\Delta q_{\mathrm{tv}} + h_{\mathrm{tv}}(\Delta q_{\mathrm{tv}})^2}{1 + \alpha\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}} \tag{31}$$

If the image quadratic function is defined using Eq. (28), then the relation between both eigenvalues is

$$\lambda = \lambda^* + \frac{\sum_{i=1,i\neq \mathrm{tv}}^{n}(2f_i\Delta q_i + h_i(\Delta q_i)^2)}{1 + \alpha\Delta\mathbf{q}^{\mathrm{T}}\Delta\mathbf{q}} \tag{32}$$

Evaluating the correct eigenvalue λ, either by Eq. (31) or Eq. (32), one can estimate the change $Q(\Delta\mathbf{q})$ of the quadratic model of the objective function using Eq. (6).

At this point it is interesting to compare the performance of both the RS-P-RFO and RS-I-RFO algo-

rithms, since the only difference is the coupling term between the max and min subspaces (see Table 1, columns RS-P-RFO and RS-I-RFO). As in previous cases, the optimization RS-I-RFO was carried out using the *Z*-matrix coordinates. In general, the RS-I- RFO algorithm presents much better performance than the RS-P-RFO algorithm, which shows numerically that it is important to take into account the couplings between the max and min subspaces during the optimization process. Using the RS-P-RFO algorithm, where the couplings are neglected, the optimization of the transition states of the Diels-Alder cycloaddition and the CH_3O_2H decomposition did not converge within 100 iterations. However, for reactions 1, 4, 11, 13, and 14 the RS-P-RFO algorithm shows a slightly better convergence than the RS-I-RFO method.

In Table 3 and without lost of generality, we show the behavior of the RS-I-RFO, algorithm for the location and optimization of the transition state for reaction 6 of Table 1. In this case the number of variables is 11. First, we observe that the inequality $v^* = \lambda^* \alpha < \min\{-h_{\text{tv}}, \min\{h_i\}_{i=1, i\neq \text{tv}}^{n}\}$ is always satisfied. Second, we see that $\lambda^* \leq \lambda$, where λ is evaluated using Eq. (31) and, consequently, using Eq. (6) we have the inequality $Q^*(\Delta\mathbf{q} \leq Q(\Delta\mathbf{q})$. This is due to the fact that $Q^*(\Delta\mathbf{q})$ is the variation energy associated with the quadratic function with minimum character and $Q(\Delta\mathbf{q})$ is the variation energy associated with the quadratic function with saddle point character. From iteration 6, the initial step lengths, $(\Delta\mathbf{q}^T\Delta\mathbf{q})^{1/2}$, are smaller than the trust radius R, so no micro-iterative process is needed. Note that in this case $\lambda^* = v^*$ since $\alpha = 1$. Finally, in the last three iterations, $\lambda^* = v^* \approx 0$, which means that they are genuine Newton-Raphson steps. Note that in this situation $r = 1.0$.

5 Summary

We have presented a RFO method (RS-RFO) such that, with the current selected eigenvector through the transformation given by Eq. (5), the corresponding displacement vector falls within the current trust region without using any type of scaling. In this way the displacement vector possesses the correct direction. Following the idea of Banerjee et al. [2] to locate saddle points, the partitioned form of the RS-RFO method (RS-P-RFO) has been defined.

Finally, we have applied the RS-RFO method to the optimization of the image quadratic function of the quadratic model to locate saddle points [21]. This method has been called RS-I-RFO. We have shown that the RS-I-RFO algorithm and the QA/TRIM and RQNR [9] algorithms are the same, so the QA/TRIM and RQNR techniques to locate saddle points can be defined as the [2/2] Padé approximant of the corresponding image quadratic function coupled with a restricted step. We also have shown that in general to optimize and locate saddle points, the RS-I-RFO method is superior to the RS-P-RFO method.

Table 3. Behavior of the RS-I-RFO algorithm in the location and optimization of the transition state of reaction 6 of Table 1

Iteration	R[a]	$(\Delta\mathbf{q}^T\Delta\mathbf{q})^{1/2}$ [b]	α[c]	λ^*[d]	v^*[e]	$h_{\min}^{f}$	λ[g]	$Q^*(\Delta\mathbf{q})$[h]	$Q(\Delta\mathbf{q})$[i]	r[j]	$(\mathbf{g}^T\mathbf{g}/n)^{1/2}$ [k]
1	0.030	0.030 (11)	9.7×10^2	–4.9	-4.8×10^3	1.0×10^2	–1.3	–4.60	–1.18	0.99	49.38
2	0.042	0.042 (10)	4.4×10^2	–6.0	-2.6×10^3	9.9×10^1	–1.7	–5.33	–1.49	0.97	42.52
3	0.060	0.060 (10)	1.8×10^2	–6.2	-1.1×10^3	9.4×10^1	–2.2	–5.15	–1.85	0.95	32.14
4	0.085	0.085 (9)	8.7×10^1	–4.9	-4.3×10^2	8.7×10^1	–3.2	–4.02	–2.60	0.94	19.37
5	0.120	0.120 (8)	3.0×10^1	–3.8	-1.1×10^2	1.1×10^2	–3.3	–2.71	–2.35	0.96	10.82
6	0.170	0.108 (0)	1.0	–1.3	–1.3	1.0×10^2	–1.3	–0.66	–0.64	0.99	3.94
7	0.170	0.005 (0)	1.0	-5.9×10^{-3}	-5.0×10^{-3}	1.0×10^2	-4.9×10^{-3}	-3.0×10^{-3}	-2.5×10^{-3}	0.90	0.57
8	0.170	5.0×10^{-4} (0)	1.0	-1.6×10^{-4}	-1.6×10^{-4}	9.0×10^1	-1.6×10^{-4}	-8.0×10^{-5}	-8.0×10^{-5}	1.00	0.12
9	0.170	1.9×10^{-4} (0)	1.0	-1.3×10^{-5}	-1.3×10^{-5}	9.3×10^1	-1.3×10^{-5}	-7.0×10^{-6}	-7.0×10^{-6}	1.00	0.03
10	0.170	1.6×10^{-4} (0)	1.0	-2.0×10^{-6}	-2.0×10^{-6}	5.1×10^1	-2.0×10^{-6}	-1.0×10^{-6}	-1.0×10^{-6}	1.00	0.01

[a] The current trust radius in Å rad
[b] The step length in Å rad. In parentheses are the number of micro-iterations needed to reach the trust radius if the initial $(\Delta\mathbf{q}^T\Delta\mathbf{q})^{1/2} > R$; see text for more details
[c] The parameter α that appears in Eq. (8d) in $Å^{-2}$ rad^{-2}
[d] The lowest eigenvalue of the I-RFO generalized eigenvalue equation that satisfies the restricted step condition; see text for more details
[e] The shift parameter defined as $v^* = \alpha\lambda^*$, which is the Lagrangian multiplier that appears in the QA/TRIM and RQNR algorithms [9]
[f] The $h_{\min} = \min\{-h_{\text{tv}}, \min\{h_i\}_{i=1, i\neq \text{tv}}^{n}\}$, where $\{h_i\}_{i=1}^{n}$ is the set of eigenvalues of the current Hessian matrix in kcal mol^{-1} $Å^{-2}$/kcal mol^{-1} rad^{-2}/kcal mol^{-1} $Å^{-1}$ rad^{-1}; see text for more details
[g] The eigenvalue of the AH corresponding to the quadratic model of the objective function evaluated using Eq. (31)
[h] The predicted quadratic variation energy of the image model function, in kcal mol^{-1}. Evaluated using Eq. (6)
[i] The predicted quadratic variation energy of the model function, in kcal mol^{-1}. Evaluated using Eqs. (31) and (6)
[j] The coefficient $r = (E_{k+1} - E_k)/Q(\Delta\mathbf{q}_k)$ at the iteration k gives the goodness of the quadratic model at this iteration; see text for more details
[k] the rms gradient norm, in kcal mol^{-1} $Å^{-1}$/kcal mol^{-1} rad^{-1}, where n is the number of variables. In this example, $n = 11$

Acknowledgements. We are indebted to Professor S. Olivella for his valuable suggestions. We acknowledge Professor H. B. Schlegel for a reprint of [16]. This research was supported by the Spanish DGICYT (Grant PB95-0278-C02-01).

References

1. Schlegel HB (1995) In: Yarkony DR (ed) Modern electronic structure theory. World Scientific, Singapore pp 459–500
2. Banerjee A, Adams N, Simons J, Shepard K (1985) J Phys Chem 89:52
3. Jensen F (1995) J Chem Phys 102:6706
4. Simons J, Jørgensen P, Taylor J, Ozment J (1983) J Phys Chem 87:2745
5. Jensen HJA, Jørgensen P (1984) J Chem Phys 80:1204
6. Fletcher R (1987) Practical methods of optimization, Wiley, New York
7. Khait YG, Panin AY, Averyanov AS (1995) Int J Quantum Chem 54:329
8. Anglada JM, Bofill JM, (1997) Int J Quantum Chem 62:153
9. (a) Helgaker T (1991) Chem Phys Lett 182:503; (b) Culot P, Dive G, Nguyen VH, Ghuysen JM, (1992) Theor Chim Acta 82:189; (c) Bofill JM (1994) J Comput Chem 15:1 (1994).
10. Shepard R, Shavitt I, Simons J (1982) J Chem Phys 76:543
11. Bell S, Crighton JS, Fletcher R, (1981) Chem Phys Lett 82:122
12. Bofill JM, Comajuan M, (1995) J Comput Chem 16:1326
13. Hebden MD (1973) AERE Harwell Report TP515
14. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902
15. Stewart JJP (1983) QCPE Bull 3:101; Olivella S (1984) QCPE Bull 4:109
16. Peng C, Schlegel HB (1994) Isr J Chem 33:449
17. Anglada JM, Bofill JM, (1998) 19:349
18. Baker J, Chan F (1996) J Comput Chem 17:888
19. Bofill JM (1996) Chem Phys Lett 260:359
20. Gill PE, Murray W, Wright MH (1981) Practical optimization. Academic, New York
21. Smith CM (1988) Theor Chim Acta 74:85; (1990) Int J Quantum Chem 37:773
22. Sun JQ, Ruedenberg K, (1994) J Chem Phys 101:2157