Optimization
Vol. 52, No. 3, June 2003, pp. 317–331

A VALLEY FOLLOWING METHOD

WOLFGANG QUAPP*

Mathematical Institute, University Leipzig, Augustus-Platz, D-04109 Leipzig, Germany

(Received 24 July 2002; In final form 8 April 2003)

We present a procedure to follow the ''path along the valley floor'' of a hypersurface. The aim is either to find minima, or to go from a minimum to a saddle point of index one, if the saddle is at the top of the valley floor. The motivation is that of taking into account local nonconvexity of the hypersurface and possibly to determine valleys. The method uses a projector technique where the projector is built by the tangent of the valley floor line. The projector is applied to the gradient and Hessian matrix of a given function, and it is used for predictor and corrector steps in path following. The resulting path is the ''valley floor gradient extremal'' which corresponds to the smallest (absolute) eigenvalue of the Hessian. Convergence properties are analysed.

Keywords: Stationary points; Path following; Projected gradient; Newton flow; Gradient extremal

Mathematics Subject Classifications 1991: Primary 90C26, 58K05; Secondary 58-04, 65H17, 53A07

1. INTRODUCTION

The so-called minimum path of theoretical chemistry [17] is roughly defined as the line which connects two minimizers by passing the saddle point of a potential surface following the valley in between. A possibility for this line is the steepest descent path from saddle point, i.e. the stable manifold w.r.t. the standard gradient flow. However, the latter is not always the valley floor [18,19]. Newton trajectories [21,22] can be used only in certain special cases for the valley floor. The gradient extremal [3], see also [12], appeared to represent a suitable ''ansatz'' for a valley floor. However, up to [23], the procedure for the calculation of the gradient extremal has required some expansive third derivatives of the surface [30]. Further, there are many additional solution curves and turning points [13,20,23]. Usually, however, one needs only the valley path from minimum to saddle point. In this article, the combination of the gradient extremal concept with the ''reduced gradient following'' (RGF) [21–25] opens a manageable way to follow the valley floor of the surface. The original RGF finds a Newton trajectory corresponding to a selected gradient direction (the search direction). There are curves which pass all stationary points in most cases (however, see [32] for a counterexample). RGF is an interesting procedure in order

*E-mail: quapp@rz.uni-leipzig.de

ISSN 0233-1934 print: ISSN 1029-4945 online © 2003 Taylor & Francis Ltd
DOI: 10.1080/0233193031000149902

to determine all types of stationary points by way of trial [21]. Now, we iteratively replace the constant search direction of RGF by the tangent of the current curve. The new method proves a practicable algorithm for searching minima, or saddle points of index one. The latter play a crucial role in global optimization, cf. [16], Remark 5.2.8. The method uses explicitely the valley structure and is able to find the saddle point if it is at the top of the valley. The valley may be of interest by itself as it is the case in theoretical chemistry or molecular spectroscopy. There it is assumed that the wave packet of a molecular vibration moves along the valley of the potential energy hypersurface of the electronic energy. If such a vibration is further excited, it can lead to a chemical reaction [31]. If only the stationary points are searched for, the procedure may be used with crude steps along the valley. The search is restricted to the region of attraction of any minimum or saddle of index one. Thus, no global search is the aim. However, if the method is combined by successive searches: minimum $\rightarrow$ saddle $\rightarrow$ next minimum $\rightarrow$ next saddle and so on, one is able to explore diverse low lying parts of the surface (cf. [16], Chapter 8).

The article is organized as follows: Section 2 shortly recalls the fundamentals of the RGF method, and Section 3 repeats the definition of gradient extremals. In Section 4 the valley following method is developed. The convergence result is expressed in Theorem 4.4. Subsequently, the success is demonstrated by an Example, and Conclusions are given.

2. FOLLOWING THE PROJECTED GRADIENT (RGF)

The function $f : \mathbb{R}^n \rightarrow \mathbb{R}^1$ is assumed to be twice-continuously differentiable. We consider the solution of the extremal problem for stationary points $\nabla f(\mathbf{x}) = 0$. By $\mathbf{g}(\mathbf{x}) := \nabla f(\mathbf{x})$ we denote the gradient vector of f. We assume that stationary points of f are nondegenerate (hence, isolated), i.e.

$$\|\mathbf{g}(\mathbf{x})\| + |\mathrm{Det}(\mathbf{H}(\mathbf{x}))| \neq \mathbf{0}, \tag{1}$$

where $\|\ldots\|$ is the Euclidean norm, and $\mathbf{H}(\mathbf{x})$ is the Hessian matrix of second derivatives of f: $\mathbf{H}$ is the Jacobian matrix of first derivatives of $\mathbf{g}$. We are interested in minima and saddle points of index one.

First, we recall the differential equation of Branin [5], see also [14–16]. It utilizes the adjoint matrix $\mathbf{A}$ of the matrix $\mathbf{H}$ of f. This is defined as $((-1)^{i+j} m_{ij})^T$ where m_{ij} is the determinant of the minor of $\mathbf{H}$ obtained by deleting from $\mathbf{H}$ the ith row and the jth column. The matrix $\mathbf{A}$ fulfils

$$\mathbf{HA} = \mathrm{Det}(\mathbf{H})\mathbf{I}_n \tag{2}$$

with the (n, n)-identity matrix $\mathbf{I}_n$.

Let $x : \mathbb{R}^1 \rightarrow \mathbb{R}^n$ be a curve, $\mathbf{x}(t)$ with parameter t.

Definition The differential equation of Branin [5] is given by

$$\frac{d\mathbf{x}}{dt} = \pm\mathbf{A}(\mathbf{x}(t))\,\mathbf{g}(\mathbf{x}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0. \tag{3}$$

The vector field (3) is a *Gradient Newton-flow*, and curves satisfying (3) are called *Newton trajectories* of the *desingularized global Newton method.*

Secondly, we recall another point of view w.r.t. Newton flows, cf. [6,21,22,24]. Let $\mathbf{r} \in \mathbb{R}^n$ be a unit vector and $\mathbf{P_r}$ the orthogonal projection on the orthogonal complement of $\mathbf{r}$. Note that $\mathbf{P_r} := \mathbf{I}_n - \mathbf{r}\mathbf{r}^{\mathbf{T}}$, a matrix of rank $(n-1)$.

Definition If the gradient $\mathbf{g}(\mathbf{x})$ fulfils

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

then this gradient is called the *reduced gradient w.r.t. the direction* $\mathbf{r}$.

Definition Let $\mathbf{x} : \mathbb{R}^1 \to \mathbb{R}^n$ be a curve and $\mathbf{x}(t)$ be a point where $\|\mathbf{g}\| \neq 0$. The curve $\mathbf{x}(t)$ is called *RGF curve w.r.t. direction* $\mathbf{r}$ if

$$\mathbf{g}(\mathbf{x}(t))/\|\mathbf{g}(\mathbf{x}(t))\| = \mathbf{r} \tag{5}$$

holds for all t.

Proposition 2.1 *The differential equation of Branin* (3) *has the same solution curve as the RGF Eq.* (5), *if the initial point* $\mathbf{x}_0$ *of* (3) *satisfies* (5). *Consequently,* $\mathbf{x}'(t)$ *of* (3) *is the tangent to the solution curve of Eq.* (5).

Proof Let $\mathbf{H}$ be nonsingular. Considering the behavior of the gradient $\mathbf{g}(\mathbf{x}(t))$ along a solution, $\mathbf{x}(t)$, we obtain with (2) and (3):

$$\frac{d\mathbf{g}}{dt} = \mathbf{H}\frac{d\mathbf{x}}{dt} = \pm\mathbf{H}\,\mathbf{A}\,\mathbf{g} = \pm\mathrm{Det}(\mathbf{H})\mathbf{g}. \tag{6}$$

The gradient $\mathbf{g}(\mathbf{x}(t))$ changes proportionally to $\mathbf{g}$ itself. Hence, the direction $\mathbf{g}/\|\mathbf{g}\|$ is invariant. ■

Definition [5] A point where $\mathbf{A}\,\mathbf{g} = \mathbf{0}$ but $\mathbf{g} \neq \mathbf{0}$ is called an *extraneous singularity* of Eq. (3).

Proposition 2.2 *Solutions of the Eq.* (4) *build RGF curves w.r.t.* $\mathbf{r}$. *They connect stationary points which differ in their index by one, if no extraneous singularity is crossed. The latter are bifurcation points of RGF curves.*

Proof The first part follows by insetting (5) into (4). The second part is given in [14] in connection with the Proposition 2.1. ■

Note that RGF is an alternate definition of Newton trajectories [6]. The tangent $\mathbf{x}'(t)$ to a curve fulfilling (4) is obtained by

$$\mathbf{0} = \frac{d}{dt}[\mathbf{P_r g}(\mathbf{x}(t))] = \mathbf{P_r}\frac{d\mathbf{g}(\mathbf{x}(t))}{dt} = \mathbf{P_r H}(\mathbf{x}(t))\mathbf{x}'(t). \tag{7}$$

In general, the search direction, $\mathbf{r}$, and the tangent, $\mathbf{x}'(t)$, to the Newton trajectory w.r.t. $\mathbf{r}$ are different. The *predictor–corrector method of RGF* [21,22] is the predictor step along the tangent $\mathbf{x}'(t)$, and Newton–Raphson steps of the corrector to search for a solution of (4).

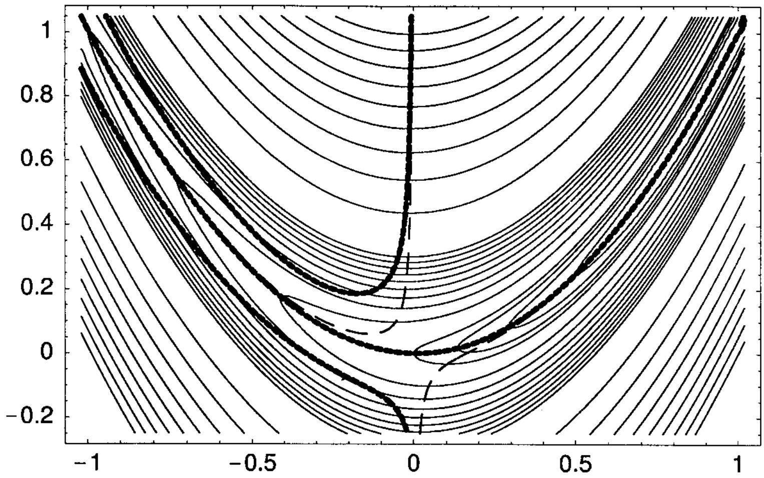

FIGURE 1 Rosenbrock surface [26] $f(x,y) = 100\ (y - x^2)^2 + (x - 1)^2$. Contour lines are at 0.25, 0.5, 0.75, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. Shown is the Newton trajectory $f_x = 0$ (dashes). Gradient extremals are dotted thick curves. The streambed GE follows perfectly the valley floor, as well as the Newton trajectory $f_y = 0$ (covered by the GE).

PROPOSITION 2.3 *Let* $\|\mathbf{g}(\mathbf{x})\| \neq \mathbf{0}$ *and* $\mathrm{Det}(\mathbf{H}(\mathbf{x})) \neq \mathbf{0}$. *The point* $\mathbf{x}$ *is the carrier of a Newton trajectory w.r.t.* $\mathbf{r}$ *where* $\mathbf{r} = \mathbf{g}(\mathbf{x})/\|\mathbf{g}(\mathbf{x})\|$.

Proof It is $\dim \ker \mathbf{P_r H(x)} = 1$. Consequently, the point $\mathbf{x}$ being a solution of (5) extends locally to a smooth manifold of dimension 1. The latter manifold is the Newton trajectory w.r.t. $\mathbf{r}$. ■

PROPOSITION 2.4 *In the neighborhood of every nondegenerate stationary point* $\mathbf{x}_S$ *starts a Newton trajectory w.r.t.* $\mathbf{r}$ *for every direction* $\mathbf{r}$.

Proof Linearization of $\mathbf{g}$ near $\mathbf{x}_S$ yields $\mathbf{g}(\mathbf{x}) \approx \mathbf{H}(\mathbf{x}_S)(\mathbf{x} - \mathbf{x}_S)$. Since $\mathbf{H}(\mathbf{x}_S)$ is nonsingular, any direction for $\mathbf{g}$ is represented near $\mathbf{x}_S$. Now, Proposition 2.3 completes the proof. ■

Figure 1 shows Newton trajectories for the Rosenbrock function [26]

$$f(x, y) = 100(y - x^2)^2 + (x - 1)^2. \tag{8}$$

The Newton trajectory w.r.t. $\mathbf{r} = (1,0)^T$, the solution of $f_y = 0$ follows the parabola $y = x^2$. It is a good approximation of the valley floor, while the trajectory w.r.t. $\mathbf{r} = (0,1)^T$, $f_x = 0$ (dashed curve), does not follow the valley throughout: it deviates at $x = 0$ from this path and it falls to two pieces there but being not a continuous model curve of the valley floor. The model demonstrates that the global Newton method (to follow a trajectory of Eq. (3)) is not able to find the minimum in all cases.

3. GRADIENT EXTREMAL (GE)

For $\mathbf{g}(\mathbf{x}) \neq \mathbf{0}$ we search for the "valley floor" of the hypersurface, f, which let be in C^3: the point showing the slowest ascent of the valley is defined by the condition that the norm of the gradient is minimized over the level set $L_c := \{\mathbf{x}/f(\mathbf{x}) = \text{constant}\}$ [3,20]. The measure for the ascent of $f(\mathbf{x})$ is the norm of the gradient

$$\sigma(\mathbf{x}) := \frac{1}{2}\|\mathbf{g}(\mathbf{x})\|^2. \tag{9}$$

Under regularity condition, the equation $f(\mathbf{x}) = c$ defines an $(n-1)$-dimensional submanifold, described by $\mathbf{x}(\mathbf{u}, c)$, where $\mathbf{u}$ is an $(n-1)$-dimensional parameter. We treat the parametric optimization problem with the objective function

$$\sigma(\mathbf{x}) \rightarrow \underset{\mathbf{x}(\,\cdot\,, c)}{\text{Min}!} \tag{10}$$

Note that both the objective function and the constraint are formed from f itself. We are interested in following a path of local minima as the parameter c changes. For almost all values of c one generally might expect that a local minimum $\mathbf{x}(c)$ of problem (10) depends differentiable on c, mainly by virtue of the implicit function theorem. The requirement for an extremal point of (10) becomes (use $\lambda(\mathbf{x})$ as Lagrange multiplier)

$$\mathbf{H}(\mathbf{x})\mathbf{g}(\mathbf{x}) = \lambda(\mathbf{x})\mathbf{g}(\mathbf{x}). \tag{11}$$

The eigenvector Eq. (11) gives use to the following definition.

Definition A point $\mathbf{x}$ belongs to the *gradient extremal* (GE) [12] if the gradient of the function f at this point is an eigenvector of the Hessian of f.

However, following a curvilinear set of consecutive GE points implies that one actually does not move in the direction of the gentlest ascent [20].

Figure 1 shows the GEs for the Rosenbrock function (8). Here, the GE to the smallest eigenvalue follows very well the valley floor. Figure 2 shows a GE which changes the character when it passes the turning points, TP. (A TP is a point where an uphill GE turns to a downhill GE, or vice versa.) But also here, one GE from Min_2 still goes to the saddle point, SP, after two sharp bends. Note that this need not be so in all cases, cf. the Fig. 15 in [20].

Definition Using the arc length s for the curve parameter, a *steepest descent* curve $\mathbf{x}(s)$ is defined by

$$\frac{d\mathbf{x}(s)}{ds} = -\frac{\mathbf{g}(\mathbf{x}(s))}{\|\mathbf{g}(\mathbf{x}(s))\|} =: -\mathbf{w}(s). \tag{12}$$

Its *curvature vector* is defined by $\mathbf{k} := (d^2\mathbf{x}/ds^2)$. A straightforward calculation shows the following alternate definition of GEs:

PROPOSITION 3.1 *Gradient extremals consist of points where steepest descent lines have zero curvature* [27]. ■

PROPOSITION 3.2 *Let $\mathbf{x}_S$ be a nondegenerate stationary point. Moreover, let the n eigenvalues of the Hessian be pairwise different. Then, exactly n GEs cross at this stationary point.*

Proof We use the Taylor series

$$f(\mathbf{x}) = f(\mathbf{x}_S) + \tfrac{1}{2}(\mathbf{x}-\mathbf{x}_S)^T\mathbf{H}(\mathbf{x}_S)(\mathbf{x}-\mathbf{x}_S) + O((\mathbf{x}-\mathbf{x}_S)^3). \tag{13}$$

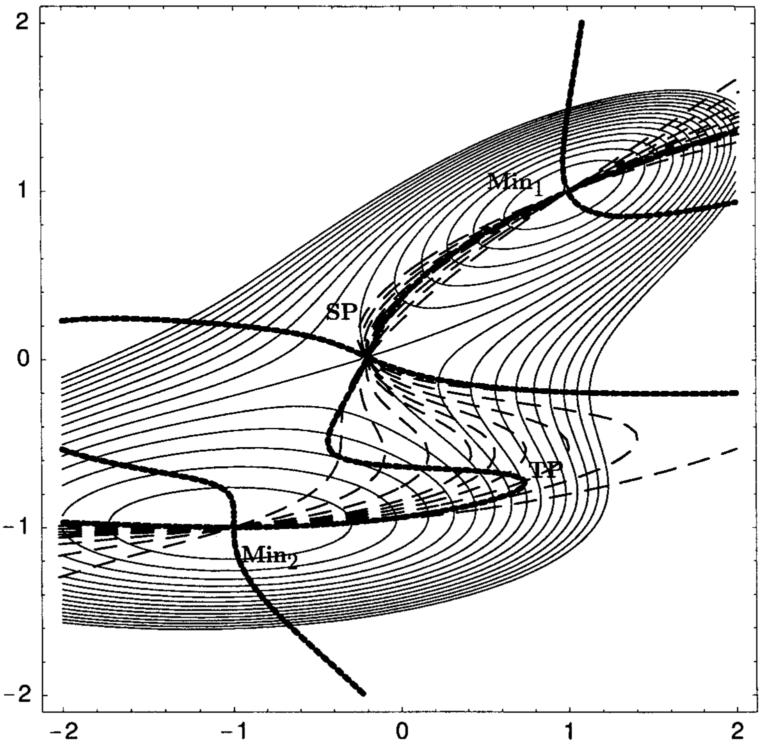

FIGURE 2 Some equidistant contour lines of the quartic surface (16). GEs are dotted thick curves, Newton trajectories are dashes. The GE from Min_1 to SP is throughout a valley GE, however, not the GE from Min_2 to SP. Its valley character ends at a turning point TP.

After affine linear coordinate transformation to normal coordinates $\mathbf{y}$, we find in Eq. (13) the representation $1/2\sum\lambda_i y_i^2$ in the second order. Because $\lambda_i \neq 0$, there are n normal coordinates y_i being the eigenvectors of $\mathbf{H}$, as well as they are the gradient directions of f, if one leaves the stationary point, $\mathbf{x}_S$, along one of these n directions. ■

Definition A *streambed gradient extremal* [28,30] is the valley GE to the smallest (absolute) eigenvalue, λ_1, of the Hessian in Eq. (11), thus $|\lambda_1| < \lambda_i$, $i = 2, \ldots, n$.

We understand this GE to be the valley floor of the hypersurface f.

Note The level L_c may not contain stationary points of $\sigma(\mathbf{x})$, e.g. if L_c is unbounded. The existence of a GE is a special event, in contrast to the Newton flow which exists through every regular point excluding some singular points [16]. Also stationary points which are in a certain common "neighborhood" may not be connected by a "direct" GE [12].

PROPOSITION 3.3 *Let $\mathbf{x}_S$ be a nondegenerate stationary point. Moreover, let the n eigenvalues of the Hessian be pairwise different. Then, each of the n GEs is tangential to a certain Newton trajectory at $\mathbf{x}_S$.*

Proof We choose the eigenvector of a given GE to be the search direction $\mathbf{r}$ of a Newton trajectory. The use of Proposition 2.4 finishes the proof. ■

We imagine to follow locally a Newton trajectory to direction $\mathbf{r}=\mathbf{Ag}/\|\mathbf{Ag}\|$.

PROPOSITION 3.4 *A point* $\mathbf{x}$ *where the gradient* $\mathbf{g}(\mathbf{x})$ *fulfils*

$$\mathbf{P}_{\mathbf{Ag}}\mathbf{g} = 0 \tag{14}$$

belongs to a gradient extremal.

Proof It is $\mathbf{P}_{\mathbf{Ag}}(\mathbf{Ag})=\mathbf{0}$, and if $\mathbf{P}_{\mathbf{Ag}}\mathbf{g}=\mathbf{0}$ then it holds that $\mathbf{g}=\alpha\,\mathbf{Ag}$ with a constant α, i.e. $\mathbf{g}=(\alpha\ \det(\mathbf{H}))\,\mathbf{H}^{-1}\mathbf{g}$ and multiplication by $\mathbf{H}$ gives $\mathbf{Hg}=(\alpha\ \det(\mathbf{H}))\mathbf{g}$. We obtain the eigenvector Eq. (11) of the GE. ■

PROPOSITION 3.5 *A point* $\mathbf{x}$ *where the tangent of a Newton trajectory is parallel to the gradient belongs to a gradient extremal.*

Proof If $\mathbf{e}_1,\ldots,\mathbf{e}_n$ are the eigenvectors of $\mathbf{H}$ with eigenvalues $\lambda_1,\ldots,\lambda_n$ then they are also the eigenvectors of the adjoint matrix, $\mathbf{A}$, but with the eigenvalues $\mu_i=\Pi_{j\neq i}\lambda_j$. This is due to the equation $\mathbf{H}\,\mathbf{e}_i=\lambda_i\,\mathbf{e}_i$, and, by multiplication with $\mathbf{A}$, we get

$$\mathbf{A}\,\mathbf{H}\,\mathbf{e}_i = \mathrm{Det}(\mathbf{H})\,\mathbf{e}_i = \lambda_i\,\mathbf{A}\,\mathbf{e}_i, \quad \text{with } \mathrm{Det}(\mathbf{H}) = \prod_{j=1}^{n}\lambda_j. \tag{15}$$

The gradient is eigenvector of $\mathbf{H}$ and of $\mathbf{A}$ on a GE. Equation (3) gives the proof. ■

Figure 2 shows the asymptotic convergence of certain Newton trajectories at three stationary points w.r.t. the valley direction expressed by the GE. A GE crosses the Newton flow at nonstationary points of f usually nontangentially. We use a quadratic surface obtained by the product of two quadratic forms

$$f(x,y) = ((x-1,y-1)H_1(x-1,y-1)^T)((x+1,y+1)H_2(x+1,y+1)^T) \tag{16}$$

with matrices being symmetric and positively definite

$$H_1 = \begin{pmatrix} 1 & -1.11 \\ -1.11 & 3 \end{pmatrix}, \quad \text{and} \quad H_2 = \begin{pmatrix} 1 & 0 \\ 0 & 12.5 \end{pmatrix}.$$

Through upper Min_1 goes a valley GE leading directly from Min_1 to the saddle. The SP is at the top of the valley. The situation changes in the lower part of Fig. 2. A valley GE is also going through Min_2, however, it ends at a TP at the slope of the surface. The next piece of the GE is not a valley line.

4. FOLLOWING THE TANGENT OF THE PREVIOUS PREDICTOR STEP (TASC)

The Method

The starting point is the predictor–corrector method of RGF. Now, we change the projector of RGF after the predictor step: the tangent direction of the previous curve point iteratively becomes the search direction used in the projector. The procedure is named the TAngent Search Concept (TASC). The aim is the asymptotic convergence of Newton trajectories to a streambed extremal. Following [23] we define the TASC step:

Assume we are at point $\mathbf{x}_k$ with $\mathbf{g}(\mathbf{x}_k)/\|\mathbf{g}(\mathbf{x}_k)\| = \mathbf{r}_k$, where $\mathbf{r}_k$ is the unit vector of the gradient of the Newton trajectory.

(i) **Predictor** Solve former Eq. (7):

$$\mathbf{P}_{\mathbf{r}_k}\mathbf{H}(\mathbf{x}_k)\mathbf{t}_k = 0, \tag{17}$$

to get the tangent direction with $\|\mathbf{t}_k\| = 1$ for the predictor step to a Newton trajectory w.r.t. $\mathbf{r}_k$, and do the step to $\mathbf{x}_k \pm s_k\mathbf{t}_k$. For a choice of the steplength s_k see later below.

(ii) **Corrector** change the search direction to $\mathbf{r}_{k+1} = \mathbf{t}_k$, compute $\mathbf{P}_{\mathbf{r}_{k+1}}$ to solve the modified equation

$$\mathbf{P}_{\mathbf{r}_{k+1}}\mathbf{g}(\mathbf{x}) = 0 \tag{18}$$

(instead of $\mathbf{P}_{\mathbf{r}_k}\mathbf{g}(\mathbf{x}) = 0$) by Newton–Raphson steps. If Eq. (18) is fulfilled then use the solution as new point $\mathbf{x}_{k+1}$.

The main idea is that (17), (18) asymptotically lead to the simultaneous equations

$$\mathbf{P}_\mathbf{r}\mathbf{H}\mathbf{r} = 0 \quad \text{and} \quad \mathbf{P}_\mathbf{r}\mathbf{g} = 0. \tag{19}$$

From (19) it follows that $\mathbf{g} = \alpha\mathbf{r}$ and $\mathbf{H}\mathbf{g} = \beta\mathbf{g}$ and, hence, the Newton trajectories corresponding to $\mathbf{g}$ coincide with a steambed extremal. In order to achieve that asymptotic behavior we prove that $\|\mathbf{P}_{\mathbf{r}_{k+1}}\,\mathbf{r}_k\| \leq C\,\beta^k$, where $0 \leq \beta < 1$. The main assumption will be an overall domination (measured by β) of the smallest absolute eigenvalue of the Hessian.

The Action of TASC

In general, the resulting curve of TASC is the valley floor:

PROPOSITION 4.1 *The TASC corrector cycle for Eq.* (18) *gives a point* $\mathbf{x}_{k+1}$ *where that gradient component is enforced which points to the direction of the eigenvector with the smallest* (*absolute*) *eigenvalue of point* $\mathbf{x}_k$.

Proof We use the Branin differential Eq. (3) which has the same solution as the method of RGF by Propositions 2.1 and 2.2. If a point of the solution curve of the RGF method w.r.t. the search direction $\mathbf{r}$ is reached, the gradient of Eq. (3) points into the same direction. Expressing $\mathbf{r}$ by the eigenvectors of $\mathbf{H}$

$$\mathbf{r} = \sum_{i=1}^{n} r_i\ \mathbf{e}_i, \tag{20}$$

we obtain with Eq. (3) the relation for the tangent direction, see Proposition 3.5,

$$\mathbf{x}' = \mathbf{A}\mathbf{r} = \sum_{i=1}^{n} r_i \left(\prod_{j \neq i} \lambda_j \right) \mathbf{e}_i. \tag{21}$$

Let λ_1 be the smallest (absolute) eigenvalue and $r_1 \neq 0$. The $\mathbf{e}_1$ component of the preceding search direction $\mathbf{r}$ is enforced, if in the next step the new direction $\mathbf{x}'$ of (21) is used, thus, if the Newton trajectory which we search for is changed. The action is the greater the larger the differences of the $\lambda_2, \ldots, \lambda_n$ are against λ_1. ■

The rate of a possible convergence of an initial direction, $\mathbf{r}$, against the path along the eigenvector $\mathbf{e}_1$ depends on the entire matrix spectrum of $\mathbf{H}$. Before we begin to deeper analyse the action of TASC and its convergence properties, we must understand in detail the

Corrector Step Let $|\lambda_1| < \lambda_2, \ldots, \lambda_n$ for the first eigenvalue of the Hessian of function f, and let the unit vector $\mathbf{e}_1$ denote the corresponding first eigenvector. Along the lines of [1,2] we start with a point $\mathbf{x}$ which fulfils

$$\mathbf{P}\mathbf{e}_1\mathbf{g}(\mathbf{x}) = 0 \tag{22}$$

at the valley floor. For simplicity, we develop the idea in the special case that the system of projector equations is fulfilled for the first normal coordinate being the direction of the valley floor, $\mathbf{e}_1$. Still more simple, system (22) can be understood as $(n-1)$ zero equations for the gradient components $i = 2, \ldots, n$ [21]. The derivative of Eq. (22) along an assumed parameter t gives the matrix $\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x})$ with

$$\dim \ker \mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x}) = 1. \tag{23}$$

The solution of (22) locally extends to a smooth one-dimensional manifold of solutions. If we assume (22) to be a system of $(n-1)$ equations, then $\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x})$ is an $(n-1,\ n)$-matrix. If we assume, again for simplicity, at point $\mathbf{x}$ normal coordinates, we have

$$\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x}) = \begin{pmatrix} 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix} \tag{24}$$

in a special diagonal form. The Moore–Penrose inverse $(\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x}))^+$ of (24) is the unique matrix with

$$\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x})(\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x}))^+ = \mathbf{I}_{n-1}. \tag{25}$$

It is the transposed matrix of (24) where additionally the eigenvalues are inverted:

$$(\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x}))^+ = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 1/\lambda_2 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & 1/\lambda_n \end{pmatrix}. \tag{26}$$

The $\ker \mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x})$ contains the searched tangent $\mathbf{t}(\mathbf{x})$. It is the vector $(1,0,\ldots,0)^T$ in normal coordinates at point $\mathbf{x}$. The reduced Hessian $\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x})$ projects any n-vector of $\mathbb{R}^n$ into an $(n-1)$-dimensional subspace $\mathbf{Y}$, where, at the other hand, we have from (26) and the choice of $\mathbf{t}$, the scalar product:

$$((\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x}))^+\mathbf{y})^T \cdot \mathbf{t}(\mathbf{x}) = 0, \quad \text{for all } \mathbf{y} \in \mathbf{Y}. \tag{27}$$

Thus, we can locally define the Newton method by

$$\begin{aligned} &\Delta \mathbf{x}^k = -(\mathbf{Pe}_1\mathbf{H}(\mathbf{x}^k))^{+}\mathbf{Pe}_1\mathbf{g}(\mathbf{x}^{\mathbf{k}}) \quad \text{and} \\ &\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta \mathbf{x}^k, \quad k = 0, 1, \ldots, \quad \text{and} \quad \mathbf{x}^0 \text{ is given on } \mathbf{t}(\mathbf{x}). \end{aligned} \tag{28}$$

So, by virtue of (27), the Newton steps are orthogonally to $\mathbf{t}(\mathbf{x})$. Their convergence properties are described [10]:

PROPOSITION 4.2 *Let* $\mathbf{Pe}_1\mathbf{g} : \mathbb{R}^n \to \mathbf{Y}$, $\mathbf{Pe}_1\mathbf{g} \in C^1(\Omega)$ *for some convex open set* $\Omega \subset \mathbb{R}^n$ *such that the derivative* $\mathbf{Pe}_1\mathbf{H}$ *of* $\mathbf{Pe}_1\mathbf{g}$ *satisfies* (23) *for all* $\mathbf{x} \in \Omega$. *Under the assumptions that at the initial point*

$$\|(\mathbf{Pe}_1\mathbf{H}(\mathbf{x}^0))^{+}\mathbf{Pe}_1\mathbf{g}(\mathbf{x}^0)\| \leq \alpha(\mathbf{x}^0), \tag{29}$$

and a Lipschitz condition at $\mathbf{x}^0$ *holds*

$$\|(\mathbf{Pe}_1\mathbf{H}(\mathbf{x}^0))^{+}[\mathbf{Pe}_1\mathbf{H}(\mathbf{x}_2) - \mathbf{Pe}_1\mathbf{H}(\mathbf{x}_3)]\| \leq \omega_0\|\mathbf{x}_2 - \mathbf{x}_3\| \tag{30}$$

for all $\mathbf{x}_2, \mathbf{x}_3 \in \Omega$ *and* $\mathbf{x}_2 - \mathbf{x}_3 \in \text{range}(\mathbf{Pe}_1\,\mathbf{H}(\mathbf{x}_1))^{+}$, *where* $\|\ldots\|$ *at the left hand side depicts a matrix norm in* $\mathbb{R}^n$, *and the proximal root condition* $h_0 = \alpha(\mathbf{x}^0)\omega_0 < 1/2$ *defines the set*

$$S(\mathbf{x}^0, r) = \{\mathbf{x} \in \mathbb{R}^n\colon \|\mathbf{x} - \mathbf{x}^0\| \leq r\} \subset \Omega, \quad \textit{for } r = \frac{1 - \sqrt{1 - 2h_0}}{h_0}\alpha(\mathbf{x}^0) \tag{31}$$

then it holds that iteration (28) *remains in* $S(\mathbf{x}^0, r)$ *and converges to some* $\mathbf{x}^*$ *with* $\mathbf{Pe}_1\mathbf{g}(\mathbf{x}^*) = 0$ *on the solution manifold. The rate of convergence is quadratic, and if*

$$R = \frac{1 + \sqrt{1 - 2h_0}}{\omega_0} \quad \textit{and} \quad \Theta = r/R \tag{32}$$

then

$$\|\mathbf{x}^k - \mathbf{x}^*\| \leq \frac{2\sqrt{1 - 2h_0}}{h_0}\frac{\Theta^{2^k}}{1 - \Theta^{2^k}}\alpha(\mathbf{x}^0). \tag{33}$$

The proof is a straightforward consequence of Theorem 4 in [7]. ■

The step length, s, of a predictor along $\mathbf{t}$ should be to choose as large as possible:

$$\mathbf{x}^0 = \mathbf{x}_0 \pm s\mathbf{t}_0. \tag{34}$$

But on the other hand, we also want the Newton method in iteration (28) to converge. That depends partly from the curvature of the solution of Eq. (18) which is described by the difference in the square brackets in (30) being equivalent to the change of the tangent $\mathbf{t}$ from one point $\mathbf{x}_0$ to the next $\mathbf{x}^*$; and to another part it depends from the curvature of the hypersurface, f, orthogonally to $\mathbf{e}_1$ which is described by the eigenvalues $\lambda_2, \ldots, \lambda_n$. Bear the special case of matrix (26) in mind, we find that the second lowest eigenvalue λ_2 has the highest influence for the estimation of ω_0 for the convergence.

With [2] we have the following:

PROPOSITION 4.3 *Consider a point* $\mathbf{x}_0$ *which fulfils* (18), *and the continuation step* (34) *with iteration scheme* (28). *Let the tangent* $\mathbf{t}(\mathbf{x}_0)$ *be a solution of* (17). *It is a vector of unit norm. Additionally to the assumptions of Proposition* 4.2 *let*

$$\|(\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x}_5))^+\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x}_4)\mathbf{t}_0\| \leq \omega_t s, \tag{35}$$

for any $\mathbf{x}_4 = \mathbf{x}_0 + s_1\mathbf{t}_0$, $\mathbf{x}_5 = \mathbf{x}_0 + s_2\mathbf{t}_0$, *and* $0 < s_1,\ s_2 \leq s$ *such that* $\mathbf{x}_4,\ \mathbf{x}_5 \in \Omega$, *and* $\mathbf{x}_0$ *is a solution of* (22). *Then the iteration* (28) *converges to a solution* $\mathbf{x}^*$ *of* (22) *for any starting point* $\mathbf{x}^0$ *from* (34) *with*

$$s \leq s_{\max} = \left(\frac{1}{\omega_0\omega_t}\right)^{1/2}. \tag{36}$$

The proof is given in [8]. ■

The Proposition 4.3 shows that a line search is possible for the predictor step of TASC. We use the line search estimation s_k of [29] assuming that the final search for stationary points is restricted to the one-dimensional line of the streambed extremal, as it is induced by Propositions 3.3 and 3.5.

Note To neglect the assumptions of Propositions 4.2 and 4.3, we anticipate a restricted step length of the corrector. Then TASC also works successfully in a generalized sense using repeated corrector steps [23]: TASC can also find saddle points in cases like in Fig. 2 if the search starts at Min_2.

The Convergence Theorem

We will show that it is possible to follow exactly the streambed GE. The discovery of the convergence resulted from computer experiments with the TASC method. Because the self consistent search of TASC around the valley GE is a surprising successful method [11], we have looked for a proof.

THEOREM 4.4 *Let* $\mathbf{x}$ *be in a convex set* $\Omega \subset \mathbb{R}^n$. *Let* $|\lambda_1| < \lambda_2, \ldots, \lambda_n$ *be for the first eigenvalue of the Hessian of function* f, *and* $\mathbf{e}_1(\mathbf{x})$ *be the corresponding eigenvector. Let the assumptions of Propositions* 4.2 *and* 4.3 *be valid. Then repeated TASC corrector steps with successive, adapted projectors* (18) *converge to the streambed gradient extremal. The smallness of* $\beta = |\lambda_1|/\lambda_2 < 1$ *is an efficiency measure: the convergence is with* β^m *where* m *is the step number.*

Sketch of the Proof Let $\lambda_1 \neq 0$. We proceed by induction.

Part (i) concerns the directions $\mathbf{r}_m$, $m = 0,1,\ldots$. Assume that $\mathbf{x}_0$ is a point on the streambed GE, then $\mathbf{g}(\mathbf{x}_0)/\|\mathbf{g}(\mathbf{x}_0)\| = \mathbf{r}_0 = \mathbf{e}_1(\mathbf{x}_0)$. The Step (34) may lead to $\mathbf{x}^0$ and the Newton iteration (28) converges to $\mathbf{x}^*$. Set $\mathbf{x}^* = \mathbf{x}_1$. We will have again $\mathbf{g}(\mathbf{x}_1)/\|\mathbf{g}(\mathbf{x}_1)\| = \mathbf{r}_0$. However, at the new point is $\mathbf{e}_1(\mathbf{x}_1) \neq \mathbf{e}_1(\mathbf{x}_0)$, in the general case, because the valley can be curvilinear. We calculate the tangent $\mathbf{t}(\mathbf{x}_1)$ of a Newton trajectory w.r.t. $\mathbf{e}_1(\mathbf{x}_0)$ from (17) by

$$\mathbf{P}\mathbf{e}_1\mathbf{H}(\mathbf{x}_1)\mathbf{t}_1 = 0 \tag{37}$$

and set $\mathbf{r}_m = \mathbf{t}_m$, $m = 1$. Construct $\mathbf{P}_{\mathbf{r}_m}$, $m = 1$, and iterate the Newton process (28) with $\mathbf{x}^0 = \mathbf{x}_m$, $m = 1$ and that projector $\mathbf{P}_{\mathbf{r}_m}$. The iteration gives again an $\mathbf{x}^* = \mathbf{x}_m$, $m = 2$.

Calculate the tangent $\mathbf{t}_m$, $m=2$, of a Newton trajectory w.r.t. $\mathbf{r}_m$, $m=1$, a.s.o.: repeat the process for $m=2,3,\ldots$ Let $\mathbf{x}_{m+1}$ be a point generated by this process. We find with (21) that the tangent of a Newton trajectory w.r.t. $\mathbf{r}_m$ points into direction $\sum_i r_{mi}(\prod_{j\neq i}\lambda_{mj})\mathbf{e}_i(\mathbf{x}_m)$, with eigenvalues λ_{mj} and eigenvectors $\mathbf{e}_i(\mathbf{x}_m)$ of the former point $\mathbf{x}_m$. Normalized, we identify the direction of the tangent with the new search direction of a Newton trajectory w.r.t.

$$\mathbf{r}_{m+1} = \frac{\sum_{i=1}^n r_{mi}/\lambda_{mi}\mathbf{e}_i(\mathbf{x}_m)}{\left\{\sum_{i=1}^n (r_{mi}/\lambda_{mi})^2\right\}^{1/2}} = \frac{\sum_{i=1}^n (|\lambda_{m1}|/\lambda_{mi})r_{mi}\mathbf{e}_i(\mathbf{x}_m)}{\left\{\sum_{i=1}^n ((\lambda_{m1}/\lambda_{mi})r_{mi})^2\right\}^{1/2}}. \tag{38}$$

It is

$$\sum_{i=2}^n r_{mi}^2 > \sum_{i=2}^n \left(\frac{\lambda_{m1}}{\lambda_{mi}} r_{mi}\right)^2, \tag{39}$$

thus if looking for the first component of the gradient, if the former $r_{m1} > 0$, it is now

$$r_{m1} = \frac{r_{m1}}{\left\{\sum_{i=1}^n r_{mi}^2\right\}^{1/2}} < \frac{r_{m1}}{\left\{\sum_{i=1}^n ((\lambda_{m1}/\lambda_{mi})r_{mi})^2\right\}^{1/2}} = \frac{r_{m1}/|\lambda_{m1}|}{\left\{\sum_{i=1}^n (r_{mi}/\lambda_{mi})^2\right\}^{1/2}} = r_{(m+1)1}. \tag{40}$$

On the left hand side we have the ratio of the first component of $\mathbf{g}(\mathbf{x}_m)$ to the first eigenvector, $\mathbf{e}_1$, at $\mathbf{x}_m$, where the right hand side gives this ratio for a tangent of a Newton trajectory w.r.t. $\mathbf{r}_m$ at point $\mathbf{x}_{m+1}$. Repeating an internal cycle of Newton steps (28) with projectors $\mathbf{P}_{\mathbf{r}_{m+k}}$ beginning with points $\mathbf{x}_{m+k}=\mathbf{x}^0$, $k=1,\ldots$, correspondingly, without doing the predictor step, leads to convergence of the first component $r_{m1}\to 1$, as well as to convergence to zero of the remaining components. They are estimated by $|r_{mi}|<c_m\beta^m$, $i=2,\ldots,n$, and c_m is bounded. This comes out by

$$|r_{(m+1)i}| = \frac{|r_{mi}||\lambda_{m1}|/\lambda_{mi}}{\left\{\sum_{i=1}^n (r_{mi}(\lambda_{m1}/\lambda_{mi})^2\right\}^{1/2}} < \frac{1}{r_{m1}}\frac{|\lambda_{m1}|}{\lambda_{mi}}|r_{mi}| < \frac{1}{r_{m1}}\beta|r_{mi}|, \quad i=2,\ldots,n. \tag{41}$$

So, the convergence is very quick if $\beta=|\lambda_1|/\lambda_2\ll 1$.

Part (ii) of the proof concerns the foot-points of the tangents $\mathbf{t}_m=\mathbf{t}(\mathbf{x}_m)$, $m=0,1,\ldots$, which we have treated in Part (i). We have to look for condition (29) in the mth iteration. We estimate the component $\|\mathbf{P}_{\mathbf{r}_{m+1}}\mathbf{g}(\mathbf{x}_{m+1})\|$ of that condition. It is $\mathbf{P}_{\mathbf{r}_{m+1}}=\mathbf{I}-\mathbf{r}_{m-1}\,\mathbf{r}_{m-1}^T$, and $\mathbf{g}(\mathbf{x}_{m+1})$ points into direction $\mathbf{r}_m$. Thus we can treat

$$\begin{aligned}\mathbf{P}_{\mathbf{r}_{m+1}}\mathbf{r}_m &= \mathbf{r}_m - \mathbf{r}_{m+1}\mathbf{r}_{m+1}^T\mathbf{r}_m \\ &= \sum_{i=1}^n \left[r_{mi}\left(1-\frac{\sum_{p=1}^n r_{mp}^2\lambda_{m1}^2/(\lambda_{mp}\,\lambda_{mi})}{\sum_{s=1}^n r_{ms}^2\lambda_{m1}^2/\lambda_{ms}^2}\right)\right]\mathbf{e}_i(\mathbf{x}_m)\end{aligned} \tag{42}$$

and

$$\|\mathbf{P}_{\mathbf{r}_{m+1}}\mathbf{r}_m\| = \left(\sum_{i=1}^n \left[\frac{r_{mi}}{\sum_{s=1}^n r_{ms}^2(\lambda_{m1}^2/\lambda_{ms}^2)}\sum_{p=1}^n r_{mp}^2\frac{|\lambda_{m1}|}{\lambda_{mp}}\left(\frac{|\lambda_{m1}|}{\lambda_{mp}}-\frac{|\lambda_{m1}|}{\lambda_{mi}}\right)\right]^2\right)^{1/2} < C\beta^m,$$

with a global constant C. The estimation leads to convergence of the points $\mathbf{x}_m$ via (33) where now $\alpha(\mathbf{x}_m)$ of the first Newton step in cycle m is estimated including the decreasing factor β^m.

Last, in **Part (iii)** we treat the special case $\lambda_1 = 0$ where the streambed line has an inflection point. The Eq. (21) comes out immediately with the new direction $\mathbf{t}_m = \mathbf{e}_1(\mathbf{x}_m)$ with $r_{m1} = 1$. This is the best case for following the $\mathbf{e}_1$-direction. ■

Leaving out from the idea of this procedure, to use a predictor–corrector scheme, the predictor step, we will really arrive the gradient direction that is parallel to the first eigenvector at a point $\mathbf{x}_m$. Then this point is situated on the streambed GE.

Note In the example below we do not use several internal cycles of Newton correctors. We adapt the streambed GE for a crude leading line only. We can use for comparison only one Newton step of the corrector per predictor step. Also along such a work-to-rule, TASC works successfully. (We have tested [11] a further improvement in the relation of predictor and corrector proposed in [6].)

5. EXAMPLE: ROSENBROCK FUNCTION

Figure 3 illustrates the action of TASC. We use the notorious Rosenbrock function (8), cf. [4]. The minimum is $f(1,1) = 0$, and the highest level is 100 in Fig. 3, but the value of $f(-1, 1)$ near the valley floor path is only 4. Along the parabola, $y = x^2$, we have $f(x, x^2) = (x-1)^2$. The parabola is the Newton trajectory $f_y = 0$, it is nearly the valley floor, which we define by the GE.

To give a general case, we start at $(-1, 0.733)$ on the "other side" of the central ridge, beside the valley floor. The first step is steepest descent to the valley ground. Searching along the valley by TASC, we take the polynomial line search [29] for the stepsize of the predictor, but a large threshold (of 0.25) for the corrector steps. This is well enough to observe one corrector step before the next predictor is done. Because the valley is curvilinear, the predictor steps are somewhat skew to the valley line, and a corrector step has to find back into the bottom region. The convergence of the solution to the minimum at $(1,1)$ is readily apparent. The eigenvalues of $\mathbf{H}$ at the minimum are $\lambda_1 = 0.4$ and $\lambda_2 = 1001.6$, thus the condition number $\lambda_{max}/\lambda_{min}$ is 2.5×10^3. For such a valley with bent valley floor (the banana valley problem) the precise determination of the minimum

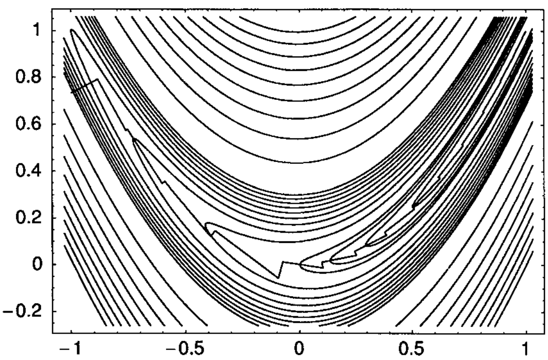

FIGURE 3 Convergence of TASC on the Rosenbrock surface [26], see caption of Fig. 1. Starting at $(-1, 0.733)$, there are 15 predictor steps to find the minimum at (1,1) along the streambed gradient extremal.

point has been rather difficult up to now. Some optimization methods converge slowly. Adapted methods give the following results [29]: Truncated Newton method: 22 iterations, #fg: 27, nonlinear conjugate gradient: 14 iterations, #fg: 31, Quasi-Newton, full-memory BFGS: 40 iterations, #fg: 47. (#fg is the number of function and gradient evaluations only; TASC uses additionally the Hessian or updates of it.) In [9] an average of 10 iterations is given for the optimization. There, the nonconvex problem (8) is converted into a convex problem by a special new variable. Examples for TASC on the Rosenbrock function for dimensions $n = 20$, and $n = 100$ are given elsewere [25].

6. CONCLUSION

TAngent Search Concept follows the streambed of a hypersurface, downhill or uphill. We demonstrate the TASC algorithm by a highly coupled problem with strong nonlinearity: the Rosenbrock function. The method performs well in practice. We use the evaluation of gradient and (updates of) the Hessian per iteration step. For minimum optimization, we can start at any point in the catchment region of a minimum and follow the gradient down the slope to the valley floor. Then we follow this line in the direction of the smallest eigenvector. For saddle search, the method can only be heuristic because TASC works if a continuous streambed exists up to the saddle. The valley floor may be defined by the gradient extremal. There is a convergence proof that this valley line is calculable as exactly as we need it. The original RGF [22] forms a tool to find minima or saddle points by Newton trajectories, where the choice of the search direction is quite arbitrary. But RGF usually diverges from the valley floor. The choice of the actual tangent in TASC overcomes the arbitrariness of the direction, and it leads to a self consistent tracing of the valley.

Acknowledgments

The work was supported by the Deutsche Forschungsgemeinschaft. The author warmly thanks D. Heidrich and M. Hirsch for many interesting discussions and A. Kripfganz for critically reading the manuscript. Thanks are due to the referees for criticism of an earlier version; especially, thanks are due to H. Th. Jongen for some deep suggestions.

Electronic Availability

The source code of TASC is implemented as a FORTRAN shell. The author would like to distribute it on request, e-mail: quapp@rz.uni-leipzig.de. It can be retrived from ⟨http://www.mathematik.uni-leipzig.de/MI/quapp/mTASC⟩

References

[1] E.L. Allgower and K. Georg (1990). *Numerical Continuation Methods – An Introduction*. Springer, Berlin.

[2] G. Bader and P. Kunkel (1992). Continuation and Gauss–Newton techniques. In: J.R. Cash and I. Gladwell (Eds.), *Computational Ordinary Differential Equations*, pp. 225–236. Claredon Press, Oxford.

[3] (a) M.V. Basilevsky and A.G. Shamov (1981). The local definition of the optimum ascent path on a multidimensional potential energy surface and its practical application for the location of saddle points. *Chem. Phys.*, **60**, 347–358; (b) M.V. Basilevsky (1981). The topography of potential energy surfaces. *Chem. Phys.*, **60**, 337–346.

[4] P. Bouvry, F. Arbab and F. Seredynski (2000). Distributed evolutionary optimization in manifold: Rosenbrock's function case study. *Inform. Sci.*, **122**, 141–159.

[5] F.H. Branin Jr. (1972). Widely convergent methods for finding multiple solutions of simultaneous nonlinear equations. *IBM J. Res. Develop.*, **16**, 504–522.

[6] I. Diener and R. Schaback (1990). An extended continuous Newton method. *J. Opt. Theory Appl.*, **67**, 87–107.

[7] P. Deuflhard and G. Heindl (1979). Affine invariant convergence theorems for Newton's method and extensions to related methods. *SIAM J. Numer. Anal.*, **16**, 1–10.

[8] P. Deuflhard, B. Fiedler and P. Kunkel (1987). Effective numerical pathfollowing beyond critical points. *SIAM J. Numer. Anal.*, **24**, 912–927.

[9] C.A. Floudas and V. Visweswaran (1993). Primal-relaxed dual global optimization approach. *J. Optimiz. Theory Appl.*, **78**, 187–225.

[10] A.A. Goldstein (1990). A modified Kantorovich inequality for the convergence of Newton's method. *Contemporary Math.*, **114**, 285–294.

[11] M. Hirsch and W. Quapp (2002). Improved RGF method to find saddle points. *J. Computat. Chem.*, **23**, 887–894.

[12] D.K. Hoffman, R.S. Nord and K. Ruedenberg (1986). Gradient extremals. *Theoret. Chim. Acta*, **69**, 265–280.

[13] F. Jensen (1995). Locating transition structures by mode following: a comparison of 6 methods on the Ar_8 Lennard-Jones potential. *J. Chem. Phys.*, **102**, 6706–6718.

[14] H. Th. Jongen, P. Jonker and F. Twilt (1987). A note on Branin's method for finding the critical points of smooth functions. In: J. Guddat, H. Th. Jongen, B. Kummer and F. Nožička (Eds.), *Parametric Optimization and Related Topics*, pp. 209–228. Akademic-Verl, Berlin.

[15] H. Th. Jongen (1990). Parametric optimization: critical points and local minima. In: E.L. Allgower and K. Georg (Eds.), *Computational Solution of Nonlinear Systems of Equations*, pp. 317–335. Lectures of Applied Mathematics, vol. 26. Amer. Math. Soc., Providence.

[16] H. Th. Jongen, P. Jonker and F. Twilt (2000). *Nonlinear Optimization in Finite Dimensions – Morse Theory, Chebychev Approximation, Transversality, Flows, Parametric Aspects*. Kluwer, Dordrecht.

[17] K. Laidler (1969). *Theory of Reaction Rates*. McGraw-Hill, New York.

[18] W. Quapp and D. Heidrich (1984). Analysis of the concept of minimum energy path on the potential energy surface of chemically reacting systems. *Theoret. Chim. Acta*, **66**, 245–260.

[19] W. Quapp (1989). Gradient extremals and valley floor bifurcation on potential energy surfaces. *Theoret. Chim. Acta*, **75**, 447–460.

[20] W. Quapp, O. Imig and D. Heidrich (1995). Gradient extremals and their relation to the minimum energy path. In: D. Heidrich (Ed.), *The Reaction Path in Chemistry: Current Approaches and Perspectives*, pp. 137–160. Kluwer, Dordrecht.

[21] W. Quapp, M. Hirsch, O. Imig and D. Heidrich (1998). Searching for saddle points of potential energy surfaces by following a reduced gradient. *J. Computat. Chem.*, **19**, 1087–1100.

[22] W. Quapp, M. Hirsch and D. Heidrich (1998). Bifurcation of reaction pathways: the set of valley-ridge-inflection points of a simple 3-dimensional potential energy surface. *Theoret. Chem. Acc.*, **100**, 285–299.

[23] W. Quapp, M. Hirsch and D. Heidrich (2000). Following the streambed reaction on potential energy surfaces: a new robust method. *Theoret. Chem. Acc.*, **105**, 145–155.

[24] W. Quapp (2001). Comment "On the quadratic reaction path evaluated in a reduced potential energy surface model and the problem to locate transition states". *J. Computat. Chem.*, **22**, 537–540.

[25] W. Quapp (2001). Searching minima of an N-dimensional surface: a robust valley following method. *Comput. Math. Appl.*, **41**, 407–414.

[26] H.H. Rosenbrock (1960). An automatic method for finding the greatest or least value of a function. *The Computer J.*, **3**, 175–184.

[27] D.J. Rowe and A. Ryman (1982). Valleys and fall lines on a Riemannian manifold. *J. Math. Phys.*, **23**, 732–735.

[28] K. Ruedenberg and J.Q. Sun (1994). Gradient fields of potential energy surfaces. *J. Chem. Phys.*, **100**, 5836–5848.

[29] (a) T. Schlick (1992). Optimization methods in computational chemistry. In: K.B. Lipkowitz and B. Boyd (Eds.), *Reviews in Computational Chemistry,* Vol. III. pp. 1–71. VCH Publishers, New York; see also (b) The Computational Science Education Project (1995). *Mathematical Optimization*, http://csepl.phy.ornl.gov/CSEP/MO/NODE11B.html.

[30] J.-Q. Sun and K. Ruedenberg (1993). Gradient extremals and steepest descent lines on potential energy surfaces. *J. Chem. Phys.*, **98**, 9707–9714.

[31] A.H. Zewail (2000). Femtochemistry: atomic scale dynamics of the chemical bond (Nobel lecture). *J. Phys. Chem.*, **104**, 5660–5694.

[32] P.J. Zufiria and R.S. Guttala (2002). On the role of singularities in Branin's method from dynamic and continuation perspectives. *Appl. Math. Computat.*, **130**, 593–618.