Array 8 (2020) 100046

Contents lists available at ScienceDirect

Array

journal homepage: www.elsevier.com/journals/array/2590-0056/open-access-journal

Narrowly avoided cusp catastrophes in Least Squares Orthogonal Distance Fitting

Alvin Penner

Fonthill, ON, Canada

ARTICLE INFO

MSC:
65D17
58K35

Keywords:
Least squares orthogonal distance fitting
Beta2-spline
Cusp catastrophe
Critical points
Topology
Streamlines
Parametric function
Transcritical bifurcation

ABSTRACT

Computer animation often requires that complex objects need to be modeled by fitting them with simpler shapes such as splines. It is necessary that these simpler shapes respond continuously as the object changes shape. This response can become complex if the topology of the error function is changing. The topological events which can occur when a spline is fit to a changing object are classified and analyzed, with a view towards defining conditions which make them possible. The Least Squares Orthogonal Distance Fitting (ODF) method is used to fit a continuous parametric function whose shape is changing. The fit is performed using a closed-shape Beta2-spline. Three types of topological events can occur: self-annihilation of two solutions, simultaneous merge of three solutions into one, and crossover of two solutions in which a local minimum and saddle point interchange roles. The events can be detected and classified by inspecting the determinants of two second-order response matrices. The last two types of event occur only within sub-manifolds defined by additional constraints not normally present in the ODF solution. The nature of these constraints is defined. The constrained events are interpreted in the context of catastrophe theory, which describes dynamical systems that can exist in multiple states with the possibility of discontinuous transitions between them. Within this theory they appear as narrowly avoided cusp catastrophes.

1. Introduction

When developing a computer animation of an object whose shape is changing, it may be necessary to first approximate the shape using a set of splines which are more easily rendered than the original object. As the object's shape changes, two types of unexpected events may occur. First, in order to produce a fit whose error is minimized, it may be necessary to undergo discontinuous changes in the fitting parameters. Second, the topology of the solution set may change due to solutions merging, crossing each other, or being created. We investigate these events using the Least Squares Orthogonal Distance Fitting method (ODF). It is typically used to fit a spline to a set of discrete shape data [1]. Here we will use it to generate solutions that undergo changes in topology when we fit a spline to a continuous two dimensional curve whose shape is changing. Three types of topological events will occur. The most common event is the annihilation of two solutions where one is a local minimum and the other is a saddle point [[2], p.63]. This can be described as a *fold catastrophe* [3]. Less common is an event in which two equivalent solutions (for example, saddle points) terminate when they meet a local minimum, leaving only a single saddle point. In this case the two equivalent solutions will have a symmetry that is twofold less than the solution they are meeting. Finally, the most unusual event is one in which a saddle point meets a local minimum, both with the same symmetry, and they interchange roles. This will be described below as a narrowly avoided *cusp catastrophe* [3,4].

The *fold catastrophe* event does not require any additional constraints to be met when it occurs: it happens simply as a result of the shape change. The last two types, however, cannot occur unless additional constraints are present. We will present two different views of these additional constraints. The first view is algebraic and takes the form of a requirement that an "augmented" second-order response matrix must be singular. The second view is geometric and takes the form of a plot of a solution path relative to a catastrophe boundary. The algebraic method is desirable because it is a precisely defined local test. However, the catastrophe theory interpretation has a qualitative advantage in displaying the relationship between multiple events: in the third type of event described above, where a local minimum and saddle point interchange roles during a narrowly avoided cusp catastrophe, we find that the interchange is always followed closely by a standard fold catastrophe in which a local minimum and saddle point coalesce and disappear. The algebraic method is not capable of linking these two events, but the catastrophe theory plot, Fig. 1, makes it qualitatively clear that these two

E-mail address: penner@vaxxine.com.

https://doi.org/10.1016/j.array.2020.100046
Received 9 June 2020; Received in revised form 26 September 2020; Accepted 21 October 2020
Available online 3 November 2020

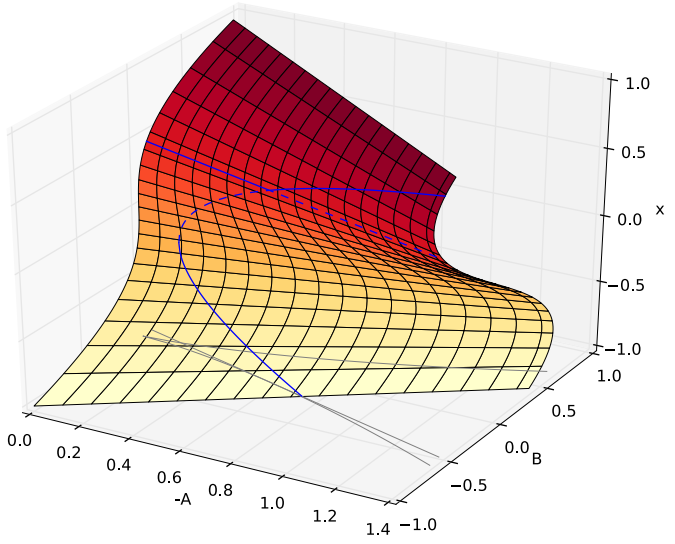


Fig. 1. Example of a cusp catastrophe, with a solution trajectory that contains both a fold catastrophe and a crossover event between stable and unstable solutions.

events are related, and that it is highly unlikely that a solution crossover could occur without the associated fold catastrophe. Fig. 1 displays a system which has two independent degrees of freedom, A and B. The solution, x, is a value that extremizes a potential energy function in dynamical systems, or in our case extremizes an error function. It may consist of either one or three values. If three solutions exist, one of them (dotted line) will be unstable, while the other two (solid lines) are stable, with the possibility of discontinuous jumps between them. The boundary between these regions is shown as a curved line when projected onto the A-B plane. The blue trajectory is a solution path that contains both a fold catastrophe and a crossover event between stable and unstable solutions, which has also been projected onto the A-B plane to show how it touches the boundary.

Events of this type are well-known in topology, where they are classified as being either "birth/death" events [5] or "interchange" events in which the critical points survive. The interchange event is significantly more complex, and has been variously described as a switch [6], swap [[7], Fig. 6], or $\mathscr{T}_{\beta}^{1}$ [[8], Fig. 3.1]. What is missing from these topological treatments is a specification of the constraints that need to be satisfied if the two critical points are to, first of all, touch, and secondly, survive and interchange. These events are also well-known in bifurcation theory [9, 10], where an "interchange" event would be termed a *transcritical bifurcation*. In this case, some examples of the constraints that need to be satisfied are given at the end of Sec. 5. The goal of the present study is to define analogous constraints in the context of ODF. We will encounter two distinct types of constraints (Secs. 6 and 7) which are sufficient to allow this to happen, in a problem which is otherwise quite simple.

The remainder of this paper is organized as follows. In Sec. 2 we derive the generic ODF method, suitable for fitting a continuous function. The family of curves we are fitting is described in Sec. 3, and the spline curve in the ODF method is specialized to be a closed-shape Beta2-spline [11]. Details of the optimization procedure are in Sec. 4, along with results. Sec. 5 defines the types of topological events that occur. Sec. 6 analyzes a symmetry-breaking event, while Sec. 7 analyzes a crossover event. These sections contain explicit definitions of the constraints that are satisfied in each case. Sec. 8 casts the above events into the language of catastrophe theory.

2. Applying ODF to a parametric function

For the ODF method we follow closely the development of [[12], Ch.2], which is a modified version of [1], to deal with the fitting of continuous curves rather than discrete data. Define $\mathbf{f}(\mathbf{a},u)=(f_x,f_y)$ as the two-dimensional spline we are adjusting to perform the fit, while $\mathbf{g}(t)=(g_x,g_y)$ is a known parametric function to be fit. $\mathbf{f}$ has adjustable parameters $\mathbf{a}$, and a parametric variable $u\in[0,N]$ where N is the number of segments, while $\mathbf{g}$ will be chosen to be a superellipse [13] with parametric variable $t\in[t_1,t_2]$. $\mathbf{g}$ has an adjustable parameter ε, which we will use to change the shape of the curve in order to see how the fitted spline responds. At each value of t we define the residual error as $||\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t)||$. In the ODF method we must adjust $u(t)$ to minimize this distance, which leads to the constraint:

$$(\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))\cdot\mathbf{f}'(\mathbf{a},u)=0. \tag{1}$$

where $\mathbf{f}'(\mathbf{a},u)=\partial\mathbf{f}/\partial u$. This simply states that the residual error vector will be perpendicular to the curve $\mathbf{f}(\mathbf{a},u)$. The error functional is

$$F(\mathbf{a},u)=\frac{1}{2}\int_{t_1}^{t_2}(\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))\cdot(\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))\,dt. \tag{2}$$

We will optimize F with respect to changes in $\mathbf{a}$ while ensuring that Eq. (1) is satisfied at all times. The optimization uses the Newton method [[14], p.348] which relies on the fact that F can be locally approximated by a quadratic form:

$$F=F_0+\mathbf{L}^t\mathbf{a}+\frac{1}{2}\mathbf{a}^t\mathbf{M}\mathbf{a} \tag{3}$$

where $\mathbf{a}^t=(a_1,\ldots,a_n)$, $\mathbf{L}^t=(L_1,\ldots,L_n)$, $L_i=dF/da_i$, and $\mathbf{M}=\{d^2F/da_i da_j\}_{i,j=1}^{n}$. The aim of the optimization is to find the critical points where $\mathbf{L}=0$, while the rate of convergence to the optimum solution will be governed by the second-order response matrix $\mathbf{M}$. The character of the solution, whether it is a minimum or saddle point, will be determined by the eigenvalues of $\mathbf{M}$. These eigenvalues are a measure of the curvature of F when moving in the direction of the corresponding eigenvector. The gradient $\mathbf{L}$ can be expressed [[12], Ch.2] as

$$L_i=\int_{t_1}^{t_2}(\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))\cdot\partial\mathbf{f}\Big/\partial a_i\,dt \tag{4}$$

while the second order response is obtained [[12], Ch.2] as $M_{ij}=\int_{t_1}^{t_2}G_{ij}(\mathbf{a},u)dt$, where

$$G_{ij}=\partial\mathbf{f}\,/\,\partial a_i\cdot\partial\mathbf{f}\,/\,\partial a_j+(\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))\cdot\partial^2\mathbf{f}\,/\,\partial a_i\partial a_j-E(u)\partial u\,/\,\partial a_i(\partial u\,/\,\partial a_j). \tag{5}$$

This contains the Hessian [[14], p.340] term $\partial^2\mathbf{f}/\partial a_i\partial a_j$, and a term in $\partial u/\partial a_i(\partial u/\partial a_j)$, which represents the contribution caused by the fact that Eq. (1) is satisfied at all times. To evaluate this term we use the relationship [[12], Ch.2]:

$$E(u)\partial u\,/\,\partial a_i=-\mathbf{f}'(\mathbf{a},u)\cdot\partial\mathbf{f}(\mathbf{a},u)\,/\,\partial a_i-(\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))\cdot\partial\mathbf{f}'(\mathbf{a},u)\,/\,\partial a_i \tag{6}$$

where

$$E(u)=\mathbf{f}'(\mathbf{a},u)\cdot\mathbf{f}'(\mathbf{a},u)+(\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))\cdot\mathbf{f}''(\mathbf{a},u). \tag{7}$$

The optimization proceeds [[12], Eq.2.9] by successively solving the equation

$$\mathbf{M}(\mathbf{a}-\mathbf{a}_0)=-\mathbf{L} \tag{8}$$

to calculate the latest increment in the vector $\mathbf{a}$, given the previous estimate $\mathbf{a}_0$. The system of equations, Eq. (1) and Eq. (8), are solved using the "variable-separation" method [1] with a tolerance of 10^{-9} in both the inner and outer loops. Further computational details are available in [[12], Ch.2].

3. Definition of spline f(a, u) and parametric curve g(t)

The function $\mathbf{g}(t)$ is taken from the family of curves in Fig. 2. These are superellipses [13]:

$$\mathbf{g}(t) = (a(\cos t)^{\varepsilon}, a(\sin t)^{\varepsilon}) \tag{9}$$

where we will set $a = 180$ to facilitate comparison with [15] for the case of a circle ($\varepsilon = 1$). The curves have a fourfold rotation symmetry and a reflection symmetry about an axis at 45°. This symmetry is denoted C_{4v} in [[16], p.60]. We will attempt to fit the curves over the range $0 < \varepsilon < 2$, which allows the curvature at some endpoints to range from zero to infinity.

When specifying $\mathbf{f}(\mathbf{a}, u)$ we need to use at least an 8-point uniform closed cubic B-spline, in order to emulate the symmetry of $\mathbf{g}(t)$. We also wish to produce a fit that is accurate to 6th order with respect to arc angle of each segment [17]. This requires that the curvature at each knot must be independently variable, which is not possible in a B-spline. We therefore generalize to a Beta2-spline, which relaxes the continuity requirements at a knot from $\mathscr{C}^2$ to $\mathscr{G}^2$ (geometric continuity) [[11], p.294]. We will decompose the Beta2-spline into eight cubic Bézier segments, and constrain the Bézier control arms to be consistent with [[11], Eq.14.3] with $\beta_1 = 1$. Note that this spline is not clamped to $\mathbf{g}(t)$ at any point, and that the slopes at each knot are not constrained. Therefore, define the parameter vector $\mathbf{a}$ using the eight Bézier endpoints and the corresponding β_2 values: $\mathbf{a} = \{x_0 - x_7, y_0 - y_7, \beta_{20} - \beta_{27}\}$.

If we define a Bézier segment i, Fig. 2, with the four control points $\{\mathbf{R}_{0i} - \mathbf{R}_{3i}\}$, then the control arm at the start point of this segment is given by $\mathbf{d}_i = \mathbf{R}_{1i} - \mathbf{R}_{0i}$. The control arms are symmetric at each node since $\beta_1 = 1$ [[11], Eq.14.2]. Therefore the control arm at the endpoint of this segment, namely $\mathbf{R}_{2i} - \mathbf{R}_{3i}$, is equal to $-\mathbf{d}_{i+1}$, so it is redundant. From [[11], Eq.14.3] we find that the components (d_{xi}, d_{yi}) of these control arms can be evaluated independently using the equation:

$$\Delta\mathbf{w} = \mathbf{C}\mathbf{d}_w \text{ for } w \in \{x, y\} \tag{10}$$

where $\Delta w_i = w_{(i+2) \bmod 8} - w_i$ for $i \in [0, 7]$. The matrix $\mathbf{C}$ is given by:

$$C_{ij} = \begin{cases} 1 & \text{for } j = i \text{ and } i \in [0, 7] \\ 4 + \beta_{2i}/2 & \text{for } j = (i+1) \bmod 8 \\ 1 & \text{for } j = (i+2) \bmod 8 \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

The relationship between $\mathbf{d}_i$ and β_{2i} is generally an inverse one: higher β_{2i} (sometimes referred to as *tension* [[11], p.307]) leads to shorter Bézier arm length $\mathbf{d}_i$. For the special case where all β_{2i} are the same, we find that $\mathbf{C}$ is singular at $\beta_{2i} = -4$. Therefore we anticipate that (roughly) $-4 < \beta_{2i} < \infty$. These are consistent with the limits highlighted in the work of de Tisi and Rossini [18]. In practice, the lowest value observed here is $\beta_2 = -3.995$ for a solution with radial symmetry, which did not lead to an abnormal termination, while the upper limit was approached on five occasions, all of which led to abnormal terminations of the solution.

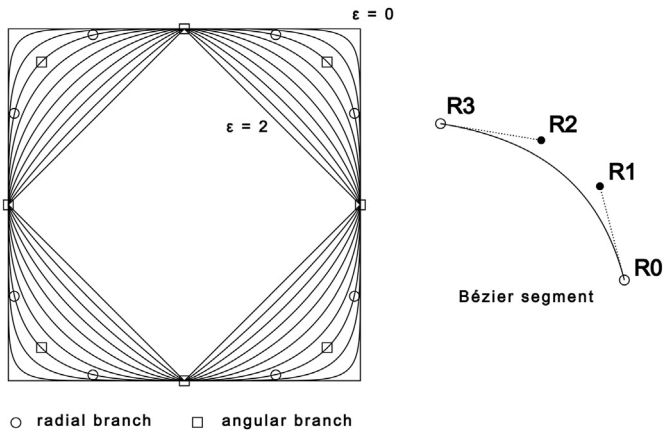


Fig. 2. A family of superellipse curves at different ε. The points are Bézier endpoints for two solution branches at $\varepsilon = 0.6$. Also shown is a typical Bézier segment with endpoints (R_0, R_3) and control points (R_1, R_2).

4. Optimization of a Beta2-Spline

The optimization requires the first-order response, Eq. (4), where $\partial \mathbf{f}(\mathbf{a}, u)/\partial a_i$ is given by the special cases:

$$df_w \Big/ dw_i = \partial f_w \Big/ \partial w_i + \sum_j (\partial f_w / \partial d_{wj}) \partial d_{wj} \Big/ \partial w_i \quad \text{for} \quad w \in \{x, y\} \tag{12}$$

$$df_w \Big/ d\beta_{2i} = \sum_j (\partial f_w / \partial d_{wj}) \partial d_{wj} \Big/ \partial \beta_{2i} \quad \text{for} \quad w \in \{x, y\}$$

where the summation includes only two Bézier control arms, $\mathbf{d}_j$, per segment. The terms $\partial f_w/\partial w_i$ and $\partial f_w/\partial d_{wj}$ are straightforward functions only of u. From Eq. (10) we obtain:

$$\partial d_{wj} \Big/ \partial w_i = C^{-1}_{j,(i-2) \bmod 8} - C^{-1}_{j,i} \text{for} \quad w \in \{x, y\} \tag{13}$$

$$\partial d_{wj} \Big/ \partial \beta_{2i} = -C^{-1}_{j,(i-1) \bmod 8} d_{wi} \Big/ 2 \text{for} \quad w \in \{x, y\}$$

The response $\partial \mathbf{f}'(\mathbf{a}, u)/\partial a_i$ in Eq. (6) is obtained directly from this by differentiation. The Hessian terms $\partial^2 \mathbf{f}/\partial a_i \partial a_j$ in Eq. (5) are zero unless they involve $\partial/\partial\beta_2$:

$$d^2 f_w \Big/ dw_i d\beta_{2j} = \sum_k (\partial f_w / \partial d_{wk}) \partial^2 d_{wk} \Big/ \partial w_i \partial \beta_{2j} \text{for} \quad w \in \{x, y\} \tag{14}$$

$$d^2 f_w \Big/ d\beta_{2i} d\beta_{2j} = \sum_k (\partial f_w / \partial d_{wk}) \partial^2 d_{wk} \Big/ \partial \beta_{2i} \partial \beta_{2j} \text{for} \quad w \in \{x, y\}$$

where the summation includes only two Bézier control arms. Application of the operator $\partial^2/\partial w_i \partial \beta_{2j}$ or $\partial^2/\partial \beta_{2i} \partial \beta_{2j}$ to Eq. (10) leads to:

$$\partial^2 d_{wk} \Big/ \partial w_i \partial \beta_{2j} = -C^{-1}_{k,(j-1) \bmod 8} \partial d_{wj} \Big/ \partial w_i \Big/ 2 \text{for} \quad w \in \{x, y\} \tag{15}$$

$$\partial^2 d_{wk} \Big/ \partial \beta_{2i} \partial \beta_{2j} = -C^{-1}_{k,(i-1) \bmod 8} \partial d_{wi} \Big/ \partial \beta_{2j} \Big/ 2 \text{for} \quad w \in \{x, y\}$$

$$-C^{-1}_{k,(j-1) \bmod 8} \partial d_{wj} \Big/ \partial \beta_{2i} \Big/ 2.$$

This completes the calculation of the optimization equations.

The results of the optimization are best displayed in radial coordinates. The radius of the Bézier endpoints is not very informative since it is always very close to the corresponding radius of $\mathbf{g}(t)$. The angle of the endpoints, however, gives useful information on symmetry. The function $\mathbf{g}(t)$ has eightfold symmetry, C_{4v}, but the solution $\mathbf{f}(\mathbf{a}, u)$ may have lower symmetry than this. If so, then the solution set will always contain multiple branches which are complements of each other with the same root-mean-square (rms) error, such that the set of all complementary branches has C_{4v} symmetry. Fig. 3 shows the variable θ_0 for the Bézier endpoint (x_0, y_0) over the range $0.6 < \varepsilon < 1.4$. The Bézier endpoints are indexed starting in the fourth quadrant, so typically $-90 < \theta_0 < 0$. The figure shows two branches that have C_{4v} symmetry: a "radial" branch with variable θ_0 ($\theta_1 = -90 - \theta_0$ and $r_0 = r_1$), and an "angular" branch with $r_0 < r_1$ and $\theta_0 = -45$ ($\theta_1 = 0$). Typical Bézier endpoints for these two branches are shown in Fig. 2 for $\varepsilon = 0.6$. These are the only two branches that have the full eightfold symmetry. In addition there are two branches with fourfold rotational symmetry, C_4. These come in pairs which are reflections of each other about an axis at 45°. They are termed "bridge" solutions because they can connect the angular and radial

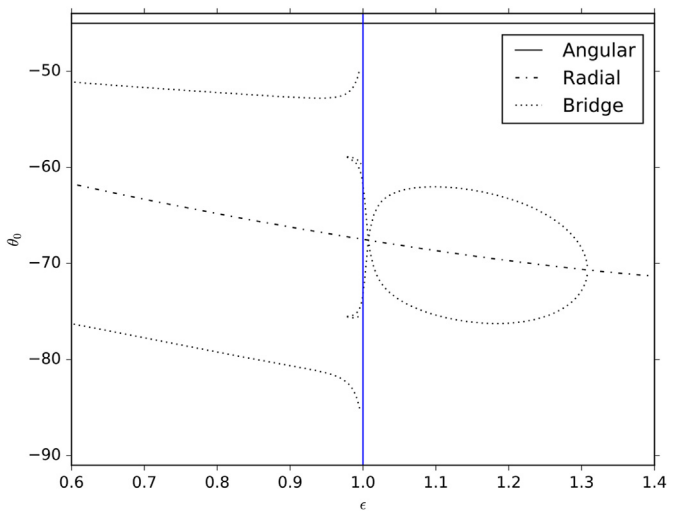


Fig. 3. Bézier end point angle θ_0 as a function of ε. Three solution branches are shown.

branches of higher symmetry. The left bridge solution, which exists only for $\varepsilon < 1$, is approaching an angular solution at $\varepsilon = 1$. The right bridge solution, which begins as a bifurcation from the radial branch at $\varepsilon = 1.308$, will touch an angular solution at $\varepsilon = 1$ in Fig. 4, and then approach a different angular branch at $\varepsilon = 1$ after reversing direction at $\varepsilon = 0.978$. These are never optimal solutions but serve an essential topological function which will be defined below.

Fig. 3 does not contain much information on the eightfold symmetric branches, other than an affirmation of the symmetry. The details of the shape are given by the β_{2i} values. Since we have nothing to compare the β_{2i} with, we choose to re-express them as Bézier control arm lengths, $d_i = |\mathbf{d}_i|$, using Eq. (10). For the special case of a circle, $\varepsilon = 1$, these arm lengths can be compared directly to [[15], Fig. 6]. There are very slight differences due to the fact that the current Beta2-spline is not clamped at the endpoints, but otherwise the two sets of solutions are qualitatively the same for the case of a circle. In Refs. [15] the "angular" solutions for the circle were subdivided into two classes: an "S-shaped" branch which includes both a symmetric ($d_0 = d_1$) solution which is a local minimum and two asymmetric branches which are also local minima, but less optimal, and an "oval-shaped" branch which contains two saddle points. This terminology is used in the current fit as well. Fig. 4 shows the arm lengths for the range of interest. For the S-shaped and Oval branches, d_0 and d_1 are distinct, for the Radial branch they are equal, and for the Bridge solutions they interchange places as we switch between the two complementary branches of the solution.

A primary goal of the current study is to analytically distinguish between the S-shaped and Oval branches, since they are qualitatively different, and since they intersect at $\varepsilon = 0.96323$. The S-shaped branch is qualitatively obtainable by fitting curvature at the Bézier endpoints, while the Oval branch is qualitatively the result of extremizing the $< y' >$ moment of the Bézier curve [15]. A secondary goal is to understand the constraints that make the bifurcation of the Radial branch at $\varepsilon = 1.308$ possible.

Fig. 5 shows the rms error. For $\varepsilon > 1$ there are five distinct unexpected terminations of either the S-shaped or Oval branches caused by $\beta_2 \to \infty$, with either d_0 or d_1 approaching zero. There is also one instance of a narrowly avoided singularity caused by $\beta_2 \to -4$ in the Radial solution at $\varepsilon = 1.69$. However, the choice of optimum solution is unambiguous for $\varepsilon > 1$. For $\varepsilon < 1$ the optimum solution path has a discontinuous jump from the S-shape branch to the Oval branch at $\varepsilon = 0.953$ followed by a continuous crossover of these two branches at $\varepsilon = 0.96323$, which will be investigated below. The Oval branch also has a *fold catastrophe* at $\varepsilon \approx 0.91$, which terminates in a manner that is consistent with [[2], p.67], namely $\Delta F \propto \sqrt{(\varepsilon - \varepsilon_0)^3}$, where ε_0 is the point where the merge occurs.

5. Two types of topological events

For each solution branch, the eigenvalues of **M** determine whether the solution is a minimum or saddle point. More generally, the Morse index [[2], p.55], which counts the number of negative eigenvalues [[19], Sec.3], will determine the multiplicity of the saddle points, if they exist. Since the object we are fitting is a closed shape, we anticipate that the number of local minima at each ε should equal the number of saddle points. Fig. 6 shows the lowest eigenvalues, eig_0, for the four types of branches. The plot often has cusps because it is quite common for different eigenvalues of the same solution to cross each other without interacting. We are interested only in the instances where an eigenvalue crosses zero or terminates at zero, which signals a topological event. The rule for these events (if higher order saddle points are involved) is that *Euler's characteristic* number [19] must be preserved [[20], Fig. 3–22, p.149]. In simplest terms this states that one local minimum must cancel one saddle point, which is the case for a fold catastrophe, or that a single saddle point must cancel a "2-saddle point" (Morse index = 2) [19]. In Fig. 6, we see that at $\varepsilon = 1.4$ the S-shape solution is a local minimum and

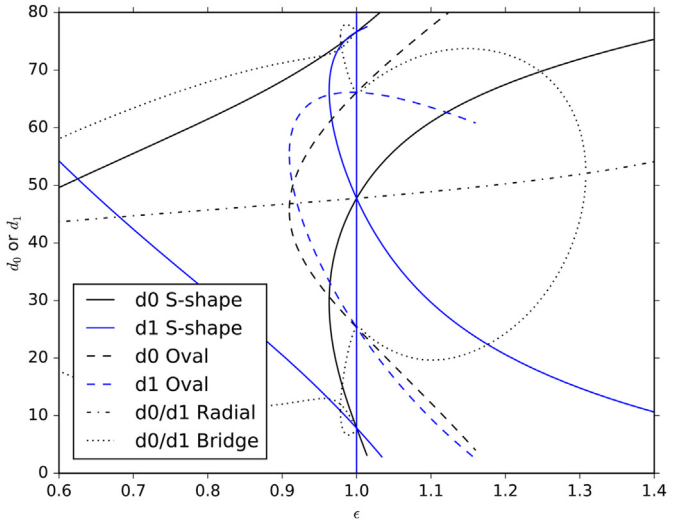


Fig. 4. Bézier control arm lengths d_0 and d_1 as a function of ε. The Angular solution branch is subdivided into S-shape and Oval branches.

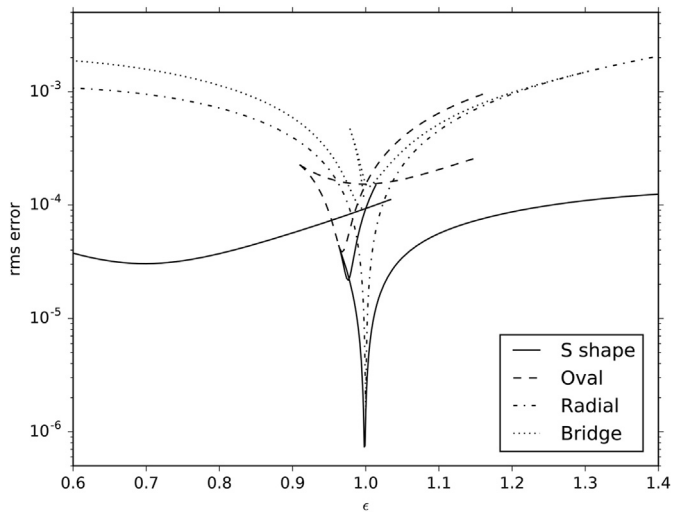


Fig. 5. Logarithm of rms error as a function of ε.

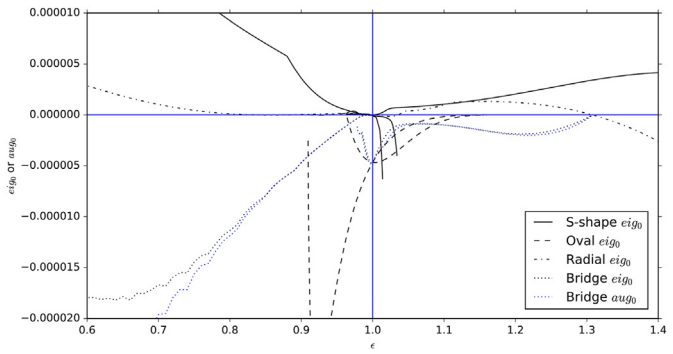


Fig. 6. Lowest eigenvalues of **M** (eig_0) and $\mathbf{M}^+$ (aug_0) as a function of ε. aug_0 is shown only for the Bridge solution, for simplicity.

the Radial solution is a saddle point, so the set of solutions is topologically balanced. However, at $\varepsilon = 1.308$ the Radial solution becomes a local minimum, so we now require two saddle points to balance this change. These are provided by the two fourfold symmetric bridge solutions for $\varepsilon < 1.308$, which serve an essential topological function even though they are not optimal fits. This event will be called a "3–1 termination" event, Sec. 6, in which a pair of solutions of lower symmetry terminates and changes the character of a higher symmetry branch. This is an example of a *pitchfork bifurcation* [[3], Fig. 7], which is normally associated with *symmetry-breaking* [[9], p.55]. It is made possible by the symmetry properties of the solutions that contribute to it [[3], Sec.10].

The left hand side of Fig. 6 shows a different type of topological event. At $\varepsilon = 0.6$ the solution set is topologically balanced, with S-shape and Radial minima, and a pair of saddle points in the Bridge solution. At $\varepsilon = 0.96323$ we encounter a "2-2 crossover" in which a saddle point in the S-shape branch crosses a local minimum in the Oval branch and interchanges character with it, Sec. 7. In this case both branches are eightfold symmetric and neither disappears. This is an example of a *transcritical bifurcation* [9,10]. Bifurcations of this type always require an additional constraint to be met. In [[9], Sec.3.2] it is the requirement that one fixed point must remain constant while another fixed point crosses over it, where the term "fixed point" refers to an equilibrium solution of a differential equation. In [[10], Fig.1.3a] it is the constraint $\alpha = \beta$, which leads to a special type of *imperfect bifurcation* (In [[3], Fig.18], this same constraint is re-expressed as $b = -a^3$.). The need for an additional constraint to explain ODF solution crossovers was also discussed in [[12], Sec.5.2]. In the context of ODF, the constraint is that two second-order response matrices are simultaneously singular, which leads to a smooth crossover of solutions rather than an avoided crossing or an abnormal termination.

6. Response to a change in shape: "3–1 termination"

The rest of this paper discusses methods of detecting such events analytically and categorizing them graphically. The question is: why are these solutions well-behaved when they meet? The matrix **M** has singularities at each event; therefore, according to Eq. (8), we should see a singularity in $d\mathbf{a}/d\varepsilon$, but Fig. 4 shows no such effect. To explain this we need to develop an analytical expression for $d\mathbf{a}/d\varepsilon$, evaluated under the constraints that $L_i = 0$ and Eq. (1) are both satisfied at all times. Following the derivation of [[12], Eq.5.5], we find:

$$\mathbf{M}(d\mathbf{a}/d\varepsilon) + \mathbf{Y} = 0 \tag{16}$$

where $Y_i = d^2F/da_i d\varepsilon = \int_{t_1}^{t_2} \mathscr{I}_i(\mathbf{a},u)dt$, and

$$\mathscr{I}_i(\mathbf{a},u) = \partial(\mathbf{f}-\mathbf{g})/\partial\varepsilon \cdot \partial\mathbf{f}/\partial a_i + (\mathbf{f}-\mathbf{g})\cdot\partial^2\mathbf{f}/\partial\varepsilon\partial a_i - E(u)\partial u/\partial\varepsilon(\partial u/\partial a_i). \tag{17}$$

This requires the calculation of $\partial u/\partial\varepsilon$, which is given by

$$E(u)\partial u/\partial\varepsilon = -\mathbf{f}'(\mathbf{a},u)\cdot\partial(\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))/\partial\varepsilon - (\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))\cdot\partial\mathbf{f}'(\mathbf{a},u)/\partial\varepsilon. \tag{18}$$

To interpret Eq. (16) we switch to a representation in which **M** is diagonalized. Define a unitary transform **U**, such that $\mathbf{U}^t\mathbf{M}\mathbf{U} = \mathbf{D}$ (diagonal). Equation (16) can be rewritten as $\mathbf{D}\mathbf{U}^t(d\mathbf{a}/d\varepsilon) + \mathbf{U}^t\mathbf{Y} = 0$. Now we see that, if a given eigenvalue, $\mathbf{D}_{00}$, is zero, and if the corresponding response, $\mathbf{U}_0^t(d\mathbf{a}/d\varepsilon)$, is to be well-behaved, then it is necessary that $\mathbf{U}_0^t\mathbf{Y} = 0$, where $\mathbf{U}_0$ is the eigenvector associated with $\mathbf{D}_{00}$.

We wish to test this assertion on the Radial branch at $\varepsilon = 1.308$. On both sides of this event the solution vector **a** shows the pattern:

$$\mathbf{a}\propto(A,B,B,A,-A,-B,-B,-A,\\ -B,-A,A,B,B,A,-A,-B,C,C,C,C,C,C,C,C)\forall\in \tag{19}$$

where, in this case, $A = x_0$ and $B = x_1$. The vector **Y** shows the same pattern, with different values for A and B. The eigenvector $\mathbf{U}_0$ shows the pattern:

$$\mathbf{U}_0\propto(E,F,-F,-E,-E,-F,F,E,\\ F,E,E,F,-F,-E,-E,-F,-G,G,-G,G,-G,G,-G,G)\forall\in \tag{20}$$

The dot product of these two vectors, $\mathbf{U}_0^t\mathbf{Y}$, is zero by symmetry, as expected. Therefore it is reasonable to expect that $d\mathbf{a}/d\varepsilon$ will be well-behaved on this branch even though the eigenvalue is going through zero. Note that this orthogonality does not exist on the lower symmetry Bridge solutions which are terminating at this point, leading to infinite slope in $d\theta_0/d\varepsilon$ in Fig. 3.

An alternate way of viewing the same issue is to define an "augmented" $\mathbf{M}^+$ matrix:

$$\begin{aligned}&\mathbf{M}^+_{ij} = \mathbf{M}_{ij} \quad \text{for} \quad \{i,j\}\in[1,n]\\ &\mathbf{M}^+_{i,n+1} = \mathbf{M}^+_{n+1,i} = Y_i = d^2F\Big/da_i d\varepsilon \quad \text{for} \quad i\in[1,n]\\ &\mathbf{M}^+_{n+1,n+1} = d^2F\Big/d\varepsilon^2\end{aligned} \tag{21}$$

where $d^2F/d\varepsilon^2 = \int_{t_1}^{t_2}\mathscr{I}(\mathbf{a},u)dt$, and

$$\mathscr{I}(\mathbf{a},u) = \partial(\mathbf{f}-\mathbf{g})\big/\partial\varepsilon\cdot\partial(\mathbf{f}-\mathbf{g})\big/\partial\varepsilon + (\mathbf{f}-\mathbf{g})\cdot\partial^2(\mathbf{f}-\mathbf{g})\big/\partial\varepsilon^2 - E(u)\partial u\big/\partial\varepsilon(\partial u/\partial\varepsilon). \tag{22}$$

The augmented matrix $\mathbf{M}^+$ is essentially the second-order response matrix obtained if we admit the variable ε as a new independent variable, comparable to $\{a_i\}$, the only difference being that F is not stationary with respect to changes in ε. It plays a role analogous to the "tangent space" ***TG*** [[10], Eq.2.2] and [[3], p.208].

Comparison of the eigenvalues of **M** and $\mathbf{M}^+$ can be informative. If $\mathbf{D}_{00}$ is an eigenvalue of **M**, and if $\mathbf{U}_0^t\mathbf{Y} = 0$, then it will also be an eigenvalue of $\mathbf{M}^+$ as well, where the augmented eigenvector is just $\mathbf{U}_0$ with an additional zero at the end. Fig. 6 shows the eigenvalue of $\mathbf{M}^+$ that is closest to zero, denoted aug_0. For the eightfold symmetric Radial branch there is always an augmented eigenvalue that agrees with eig_0 at all ε. This is the reason why this branch is able to have zeros in eig_0 while still being well-behaved. For the Bridge solution this is not normally the case, although the figure shows that eig_0 and aug_0 both approach zero at $\varepsilon = 1.308$.

The augmented matrix has been previously defined in a similar context [[2], p.66]. In that discussion it was stated that **M** and $\mathbf{M}^+$ cannot simultaneously be singular because there are not sufficient degrees of freedom to allow this to happen [[2], p.60, Assumption A]. Normally, of course, this is true: we have only one degree of freedom, ε, so we cannot simultaneously satisfy two constraints. Therefore the only way for this to be possible is if there are additional constraints in operation such that both matrices can be made to be singular on a sub-manifold of the solution space. (A similar statement is made in Ref [[10], Eq.2.10]. with reference to a bifurcation point [[10], Fig.2.1a]). In the present example,

"3–1 termination", the additional constraint is symmetry, which allows symmetric solutions to intersect with asymmetric solutions without pathological behavior. In the next section an alternative constraint is described which also allows two solution branches to cross each other and interchange roles without singular behavior. This allows [[2], p.60, Assumption A] to be violated due to the presence of a newly defined sub-manifold.

7. Response to a change in shape: "2-2 crossover"

This event consists of an intersection between the S-shaped and Oval branches in Fig. 4 at $\varepsilon = 0.96323$. At a first glance, it appears that three branches (two minima plus a saddle point) have simultaneously coalesced to produce a single minimum. Upon closer inspection we find (Fig. 7a) that there are two distinct events, both of which involve only two branches. (Note the similarity of Fig. 7a to [[10], Fig.1.3a] and to [[3], Fig.18] for the intermediate cases where $b = -a^3$). At $\varepsilon = 0.9631$ a local minimum and a saddle point coalesce and disappear in a fold catastrophe. At this point $eig_0 = 0$ and the slope in Fig. 7a is infinite. At $\varepsilon = 0.96323$ the two branches cross each other at finite slope in Fig. 7a with an interchange of character from minimum to saddle point and vice-versa (This was previously termed a Type 2 crossover in [[12], Chap.5]. It has also been called an "interchange event" in Ref. [[7], Fig. 6], "height switch" in Ref. [6], and $\mathscr{T}^1_\beta$ in Ref. [[8], Fig. 3.1].). In this case both $\mathbf{M}$ and $\mathbf{M}^+$ are singular. As before, the simultaneous singularity of these two matrices is a necessary condition for $d\mathbf{a}/d\varepsilon$ to be well-behaved. There are two questions that need to be addressed: where do we find the necessary degrees of freedom to make $\mathbf{M}$ and $\mathbf{M}^+$ simultaneously singular, and are these two events related (could a "2-2 crossover" occur without a corresponding fold catastrophe)? The first question can be answered algebraically, while the answer to the second question is most easily demonstrated using catastrophe theory (Sec. 8).

As noted in Sec. 4, the S-shape branch is a "typical" solution in that it is qualitatively comparable to fitting curvature at the endpoints. The oval-shaped solution is abnormal since it is related to an extremum of the $< y' >$ moment. Upon closer inspection, we find that the oval-shaped solution yields an unexpected relationship, which is not true on the S-shape branch: the variables $(\mathbf{f}(\mathbf{a},u) - \mathbf{g}(t)) \cdot \partial\mathbf{f}/\partial a_i$ are linearly dependent on each other $\forall t$. Combining this with Eq. (1) we obtain the more general expression

$$\mathbf{f}'(\mathbf{a},u) = h(u) \sum_{i=1}^{n} \mu_i \partial\mathbf{f} \Big/ \partial a_i \tag{23}$$

in which $h(u)$ is to be determined and μ_i are functions of ε but not u. The calculation of μ_i will be described below (Sec. 7.1). For now we note an interesting consequence of Eq. (23). From Eq. (4) and Eq. (23) we have:

$$\begin{aligned} \sum_{i=1}^{n} \mu_i L_i &= \int_{t_1}^{t_2} (\mathbf{f}(\mathbf{a},u) - \mathbf{g}(t)) \cdot \sum_{i=1}^{n} \mu_i \partial\mathbf{f} \Big/ \partial a_i dt \\ &= \int_{t_1}^{t_2} (\mathbf{f}(\mathbf{a},u) - \mathbf{g}(t)) \cdot \frac{\mathbf{f}'(\mathbf{a},u)}{h(u)} dt \\ &= 0 \end{aligned} \tag{24}$$

where we have used Eq. (1). This equation represents a new constraint on the $\{L_i\}$: this releases one of the original constraints we would normally impose on $\mathbf{L}$, which gives us a new degree of freedom we may be able to use to set the determinant of $\mathbf{M}^+$ to zero. In effect, it defines a sub-manifold within which a crossing of solutions becomes theoretically possible. We now need to derive appropriate conditions for this to occur.

The function $h(u)$ can be calculated as follows. Taking a derivative of Eq. (23) we find:

$$\begin{aligned} \mathbf{f}''(\mathbf{a},u) &= h(u) \sum_{i=1}^{n} \mu_i \partial\mathbf{f}' \Big/ \partial a_i + h'(u) \sum_{i=1}^{n} \mu_i \partial\mathbf{f} \Big/ \partial a_i \\ &= h(u) \sum_{i=1}^{n} \mu_i \partial\mathbf{f}' \Big/ \partial a_i + \frac{h'(u)}{h(u)} \mathbf{f}'(\mathbf{a},u). \end{aligned} \tag{25}$$

Substitute this into Eq. (7) to obtain:

$$E(u) = \mathbf{f}'(\mathbf{a},u) \cdot \mathbf{f}'(\mathbf{a},u) + h(u)(\mathbf{f}(\mathbf{a},u) - \mathbf{g}(t)) \cdot \sum_{i=1}^{n} \mu_i \partial\mathbf{f}' \Big/ \partial a_i \tag{26}$$

where we have used Eq. (1) to eliminate the term $h'(u)(\mathbf{f}(\mathbf{a},u) - \mathbf{g}(t)) \cdot \mathbf{f}'(\mathbf{a},u)/h(u)$, which is zero. Now operate on Eq. (6) to obtain:

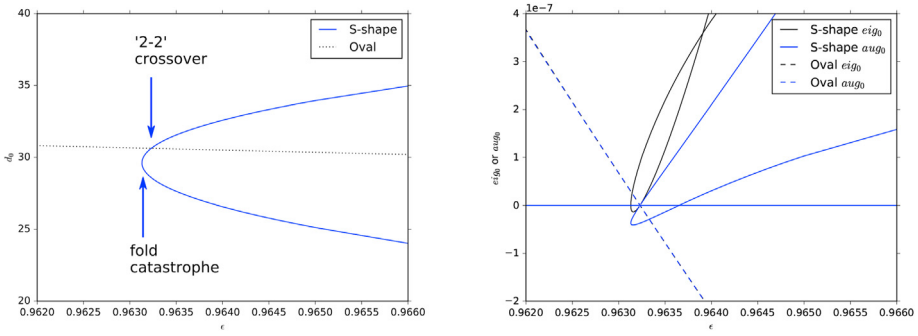
$$\begin{aligned} E(u) \sum_{i=1}^{n} \mu_i \partial u \Big/ \partial a_i &= -\mathbf{f}'(\mathbf{a},u) \cdot \sum_{i=1}^{n} \mu_i \partial\mathbf{f} \Big/ \partial a_i - (\mathbf{f}(\mathbf{a},u) - \mathbf{g}(t)) \cdot \sum_{i=1}^{n} \mu_i \partial\mathbf{f}' \Big/ \partial a_i \\ &= -\mathbf{f}'(\mathbf{a},u) \cdot \mathbf{f}'(\mathbf{a},u) \Big/ h(u) - (\mathbf{f}(\mathbf{a},u) - \mathbf{g}(t)) \cdot \sum_{i=1}^{n} \mu_i \partial\mathbf{f}' \Big/ \partial a_i. \end{aligned} \tag{27}$$

Comparing this to Eq. (26) we find:

$$h(u) \sum_{i=1}^{n} \mu_i \partial u \Big/ \partial a_i = -1 \tag{28}$$

where $h(u)$ can be viewed as a scalar normalization function in Eq. (23). If we eliminate $h(u)$ from these equations we can obtain a further interesting constraint. Substitute Eq. (28) into Eq. (23) to obtain:

$$\mathbf{f}'(\mathbf{a},u) = -\sum_{i=1}^{n} \mu_i \partial\mathbf{f} \Big/ \partial a_i \Big/ \sum_{i=1}^{n} \mu_i \partial u \,/\, \partial a_i \tag{29}$$



(a) arm length d_1 (b) eigenvalues eig_0 and aug_0

Fig. 7. Solution properties versus ε during a crossover event between the Oval and S-shaped branches. Also shown is the associated fold catastrophe event.

$$\mathbf{f}'(\mathbf{a},u)\sum_{i=1}^{n}\mu_i\partial u\Big/\partial a_i+\sum_{i=1}^{n}\mu_i\partial\mathbf{f}\Big/\partial a_i=0 \tag{30}$$

$$\sum_{i=1}^{n}\mu_i[\partial\mathbf{f}/\partial a_i+\mathbf{f}'(\mathbf{a},u)\partial u/\partial a_i]=0 \tag{31}$$

which represents a new constraint on the total response of $\mathbf{f}(\mathbf{a},u)$ with respect to $\mathbf{a}$, which is valid only on the oval-shaped branch. We note in passing that analogous constraints have also been found to be satisfied on oval-shaped branches of the ODF solutions obtained when fitting a clamped cubic and/or quartic Bézier to a 45° segment of $\mathbf{g}(t)$, where $\mathbf{g}(t)$ is a hypoTrochoid shape with C_{4v} symmetry [[12], Chaps.4, 8]. In this case they are also associated with a "2-2 crossover".

7.1. Implications for $\boldsymbol{M}$

In order to define this class explicitly, we need to calculate $\{\mu_i\}$. Begin with Eq. (23) and express it as a parallel relationship:

$$\mathbf{f}'(\mathbf{a},u)\Big\|\sum_{i=1}^{4}\mu_i\partial\mathbf{f}\Big/\partial a_i$$
$$\frac{f_y'}{f_x'}=\frac{\sum_{i=1}^{4}\mu_i\partial f_y/\partial a_i}{\sum_{i=1}^{4}\mu_i\partial f_x/\partial a_i}, \tag{32}$$

where the sum extends over all the independent variables a_i. Here we will make use of the eightfold symmetry of the Oval solution to simplify the representation. Define the Bézier control arms for one segment as in Ref. [[15], Fig. 1]:

$$\begin{aligned}\mathbf{R}_1&=(X_1,Y_1)=(X_0-d_1\sin\theta_1,Y_0+d_1\cos\theta_1)\\ R_2&=(X_2,Y_2)=(X_3+d_2\sin\theta_2,Y_3-d_2\cos\theta_2)\end{aligned} \tag{33}$$

where $\mathbf{R}_0$ and $\mathbf{R}_3$ are the endpoints. The angles θ_1 and θ_2, of $\mathbf{R}_0$ and $\mathbf{R}_3$, are fixed by symmetry, so the independent variables are $\mathbf{a}=(d_1,d_2,R_0,R_3)$. Express $\mathbf{f}(\mathbf{a},u)$ in terms of Bernstein polynomials [[11], p.224], and re-express Eq. (32) as

$$\begin{aligned}&\frac{Y_0B_{03}'+(Y_0+d_1\cos\theta_1)B_{13}'+Y_3B_{33}'+(Y_3-d_2\cos\theta_2)B_{23}'}{X_0B_{03}'+(X_0-d_1\sin\theta_1)B_{13}'+X_3B_{33}'+(X_3+d_2\sin\theta_2)B_{23}'}\\&=\frac{\mu_1B_{13}\cos\theta_1+\mu_3(B_{03}+B_{13})\sin\theta_1-\mu_2B_{23}\cos\theta_2+\mu_4(B_{23}+B_{33})\sin\theta_2}{-\mu_1B_{13}\sin\theta_1+\mu_3(B_{03}+B_{13})\cos\theta_1+\mu_2B_{23}\sin\theta_2+\mu_4(B_{23}+B_{33})\cos\theta_2}.\end{aligned} \tag{34}$$

Use the relationship $B'_{i,n}=n(B_{i-1,n-1}-B_{i,n-1})$ and rearrange:

$$\frac{d_1\cos\theta_1B_{02}+d_2\cos\theta_2B_{22}+\Delta YB_{12}}{-d_1\sin\theta_1B_{02}-d_2\sin\theta_2B_{22}+\Delta XB_{12}}=\frac{\mu_3\sin\theta_1B_{03}+\mu_4\sin\theta_2B_{33}+(\mu_1\cos\theta_1+\mu_3\sin\theta_1)B_{13}+(-\mu_2\cos\theta_2+\mu_4\sin\theta_2)B_{23}}{\mu_3\cos\theta_1B_{03}+\mu_4\cos\theta_2B_{33}+(-\mu_1\sin\theta_1+\mu_3\cos\theta_1)B_{13}+(\mu_2\sin\theta_2+\mu_4\cos\theta_2)B_{23}} \tag{35}$$

where

$$\begin{aligned}\Delta X&=X_3-X_0+d_1\sin\theta_1+d_2\sin\theta_2\\ \Delta Y&=Y_3-Y_0-d_1\cos\theta_1-d_2\cos\theta_2.\end{aligned} \tag{36}$$

Now multiply through by both denominators and organize the terms according to their use of the functions $u^i(1-u)^j$ where $i+j=5$. We find that the term in $(1-u)^5$, obtained by multiplying $B_{02}B_{03}$, is satisfied only if $\mu_3=0$. Similarly, the term in u^5 requires $\mu_4=0$. Substituting these two results in, the remaining terms are equated below:

factor	left hand side	right hand side
$u(1-u)^4$	$-\mu_1d_1\sin\theta_1\cos\theta_1$	$-\mu_1d_1\sin\theta_1\cos\theta_1$
$u^2(1-u)^3$	$\mu_2d_1\sin\theta_2\cos\theta_1-2\mu_1\Delta Y\sin\theta_1$	$\mu_2d_1\cos\theta_2\sin\theta_1+2\mu_1\Delta X\cos\theta_1$
$u^3(1-u)^2$	$-\mu_1d_2\sin\theta_1\cos\theta_2+2\mu_2\Delta Y\sin\theta_2$	$-\mu_1d_2\cos\theta_1\sin\theta_2-2\mu_2\Delta X\cos\theta_2$
$u^4(1-u)$	$\mu_2d_2\sin\theta_2\cos\theta_2$	$\mu_2d_2\sin\theta_2\cos\theta_2$

The first and last rows are satisfied trivially, which leaves only two constraints on $\{\mu_i\}$. These can be expressed as:

$$\begin{pmatrix}-2\Delta_1-2d_2\sin(\theta_2-\theta_1) & d_1\sin(\theta_2-\theta_1)\\ d_2\sin(\theta_2-\theta_1) & 2\Delta_2-2d_1\sin(\theta_2-\theta_1)\end{pmatrix}\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}=0 \tag{37}$$

where $\Delta_i=(X_3-X_0)\cos\theta_i+(Y_3-Y_0)\sin\theta_i$. We find, as expected, that this matrix is singular at all times on the oval-shaped branch in Fig. 7, so the $\{\mu_i\}$ can be calculated to within a normalization constant. Since this constraint does not involve μ_3 or μ_4, we also find that it is equally applicable to the case of a clamped cubic Bézier segment [[12], Chap.5].

Given $\{\mu_i\}$, we wish to work out the consequences for $\mathbf{M}$. Beginning with Eq. (5), and noting that the Hessian terms are zero within this particular symmetric representation, we have

$$\begin{aligned}\sum_{j=1}^{4}\mu_jG_{ij}&=\partial\mathbf{f}\Big/\partial a_i\cdot\sum_{j=1}^{4}\mu_j\partial\mathbf{f}\Big/\partial a_j-E(u)\partial u\Big/\partial a_i\sum_{j=1}^{4}\mu_j\partial u\Big/\partial a_j\\&=\frac{1}{h(u)}\partial\mathbf{f}\Big/\partial a_i\cdot\mathbf{f}'(\mathbf{a},u)+\frac{1}{h(u)}E(u)\partial u\Big/\partial a_i\end{aligned} \tag{38}$$

where we have used Eq. (23) and Eq. (28). Now use Eq. (6) to eliminate $E(u)$ and integrate G_{ij} to obtain:

$$\sum_{j=1}^{4}\mu_jM_{ij}=-\int_{t_1}^{t_2}\frac{1}{h(u)}(\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))\cdot\partial\mathbf{f}'\Big/\partial a_i dt\,\forall i \tag{39}$$

These equations provide a new mechanism by which $\mathbf{M}$ may become singular, one which is unique to the oval-shaped branch: by setting $\sum_{j=1}^{4}\mu_jM_{ij}=0\ \forall i$. We have confirmed computationally that these zeros in Eq. (39) occur $\forall i$ during a "2-2 crossover", leading to a singular $\mathbf{M}$.

7.2. Implications for $\mathbf{M}^+$

By itself, a singular $\mathbf{M}$ is not a surprising event; what is of interest is whether the $\mathbf{M}^+$ matrix will simultaneously become singular. From Eq. (17), and noting that the Hessian term is zero, we find:

$$\begin{aligned}\sum_{i=1}^{4}\mu_i\mathscr{I}_i(\mathbf{a},u)&=\partial(\mathbf{f}-\mathbf{g})\Big/\partial\varepsilon\cdot\sum_{i=1}^{4}\mu_i\partial\mathbf{f}\Big/\partial a_i-E(u)\partial u\Big/\partial\varepsilon\sum_{i=1}^{4}\mu_i\partial u\Big/\partial a_i\\&=\frac{1}{h(u)}\partial(\mathbf{f}-\mathbf{g})\Big/\partial\varepsilon\cdot\mathbf{f}'(\mathbf{a},u)+\frac{1}{h(u)}E(u)\partial u\Big/\partial\varepsilon\end{aligned} \tag{40}$$

where we have used Eq. (23) and Eq. (28). Now use Eq. (18) to eliminate $E(u)$ and integrate $\mathscr{I}_i(\mathbf{a},u)$ to obtain:

$$\sum_{i=1}^{4}\mu_i d^2F \Big/ da_i d\varepsilon = -\int_{t_1}^{t_2}\frac{1}{h(u)}(\mathbf{f}(\mathbf{a},u)-\mathbf{g}(t))\cdot\partial\mathbf{f}' \Big/ \partial\varepsilon dt. \tag{41}$$

We wish to determine whether the right hand side of Eq. (41) will go to zero simultaneously with the right hand sides of Eq. (39) $\forall i$. This would mean that a singular **M**, as defined by Eq. (39), would also lead to a singular $\mathbf{M}^+$. For the present, unclamped, spline this is trivially true, since $\partial\mathbf{f}'/\partial\varepsilon = 0$ by definition. For a clamped spline [15], a proof is required. In this case a sufficient condition for this to occur is:

$$\mathbf{f}'(\mathbf{a},u) \parallel \nu_0\partial\mathbf{f}' \Big/ \partial\varepsilon + \sum_{i=1}^{2}\nu_i\partial\mathbf{f}' \Big/ \partial a_i \forall u \tag{42}$$

where $\{\nu_i\}$ are constant with respect to u, and we have used the fact that Eq. (1) is satisfied $\forall u$. The sum extends over $\{d_1, d_2\}$. We now need to relate $\{\nu_i\}$ to $\{\mu_i\}$, to confirm that they exist. Compare this equation to Eq. (32) to obtain:

$$\nu_0\partial\mathbf{f}' \Big/ \partial\varepsilon + \sum_{i=1}^{2}\nu_i\partial\mathbf{f}' \Big/ \partial a_i \parallel \sum_{i=1}^{2}\mu_i\partial\mathbf{f} \Big/ \partial a_i. \tag{43}$$

Express $\mathbf{f}(\mathbf{a},u)$ in terms of Bernstein polynomials and Eq. (33):

$$\frac{\begin{array}{c}\nu_0\left[\left(B_{03}'+B_{13}'\right)\sin\theta_1-\left(B_{23}'+B_{33}'\right)\sin\theta_2\right]\\ +\nu_1 B_{13}{}'\cos\theta_1-\nu_2 B_{23}{}'\cos\theta_2\end{array}}{\begin{array}{c}\nu_0\left[\left(B_{03}'+B_{13}'\right)\cos\theta_1-\left(B_{23}'+B_{33}'\right)\cos\theta_2\right]\\ -\nu_1 B_{13}{}'\sin\theta_1+\nu_2 B_{23}{}'\sin\theta_2\end{array}}=\frac{\mu_1 B_{13}\cos\theta_1-\mu_2 B_{23}\cos\theta_2}{-\mu_1 B_{13}\sin\theta_1+\mu_2 B_{23}\sin\theta_2} \tag{44}$$

where we have used the definition of the hypoTrochoid $\mathbf{g}(t)$ in Ref. [15] to calculate the response $(\partial\mathbf{f}'/\partial\varepsilon)$ of the endpoints to a change in ε, which is called c in Ref. [15]. Now use the relationship $B'_{i,n} = n(B_{i-1,n-1} - B_{i,n-1})$:

$$\frac{\begin{array}{c}-\nu_0 B_{12}(\sin\theta_1+\sin\theta_2)\\ +\nu_1(B_{02}-B_{12})\cos\theta_1-\nu_2(B_{12}-B_{22})\cos\theta_2\end{array}}{\begin{array}{c}-\nu_0 B_{12}(\cos\theta_1+\cos\theta_2)\\ -\nu_1(B_{02}-B_{12})\sin\theta_1+\nu_2(B_{12}-B_{22})\sin\theta_2\end{array}}=\frac{(1-u)\mu_1\cos\theta_1-u\mu_2\cos\theta_2}{-(1-u)\mu_1\sin\theta_1+u\mu_2\sin\theta_2} \tag{45}$$

where we have canceled a common factor of $u(1-u)$ on the right hand side. Multiply through by both denominators and organize the terms according to their use of the functions $u^i(1-u)^j$ where $i+j=3$:

factor	left hand side	right hand side
$(1-u)^3$	$-\mu_1\nu_1\sin\theta_1\cos\theta_1$	$-\mu_1\nu_1\sin\theta_1\cos\theta_1$
$u(1-u)^2$	$2\mu_1\nu_0\sin\theta_1(\sin\theta_1+\sin\theta_2)$ $+2\mu_1\nu_1\sin\theta_1\cos\theta_1$ $+\mu_2\nu_1\cos\theta_1\sin\theta_2$ $+2\mu_1\nu_2\sin\theta_1\cos\theta_2$	$-2\mu_1\nu_0\cos\theta_1(\cos\theta_1+\cos\theta_2)$ $+2\mu_1\nu_1\sin\theta_1\cos\theta_1$ $+\mu_2\nu_1\sin\theta_1\cos\theta_2$ $+2\mu_1\nu_2\cos\theta_1\sin\theta_2$
$u^2(1-u)$	$-2\mu_2\nu_0\sin\theta_2(\sin\theta_1+\sin\theta_2)$ $-2\mu_2\nu_1\cos\theta_1\sin\theta_2$ $-\mu_1\nu_2\sin\theta_1\cos\theta_2$ $-2\mu_2\nu_2\sin\theta_2\cos\theta_2$	$2\mu_2\nu_0\cos\theta_2(\cos\theta_1+\cos\theta_2)$ $-2\mu_2\nu_1\sin\theta_1\cos\theta_2$ $-\mu_1\nu_2\cos\theta_1\sin\theta_2$ $-2\mu_2\nu_2\sin\theta_2\cos\theta_2$
u^3	$\mu_2\nu_2\sin\theta_2\cos\theta_2$	$\mu_2\nu_2\sin\theta_2\cos\theta_2$

The first and last rows are satisfied trivially, which leaves only two constraints on $\{\nu_i\}$. These can be expressed in matrix form:

$$\begin{pmatrix}(2\mu_1-4\mu_2)[1+\cos(\theta_2-\theta_1)] & -3\mu_2\sin(\theta_2-\theta_1) & 0\\ (4\mu_1-2\mu_2)[1+\cos(\theta_2-\theta_1)] & 0 & -3\mu_1\sin(\theta_2-\theta_1)\end{pmatrix}\begin{pmatrix}\nu_0\\ \nu_1\\ \nu_2\end{pmatrix}=0. \tag{46}$$

These constraints can always be satisfied to within a normalization constant: for example, we find ν_1/ν_0 is a linear function of μ_1/μ_2. Therefore we conclude that $\mathbf{M}^+$ may become singular simultaneously with **M**, and that this is possible only on the sub-manifold represented by the oval-shaped solution in Fig. 7.

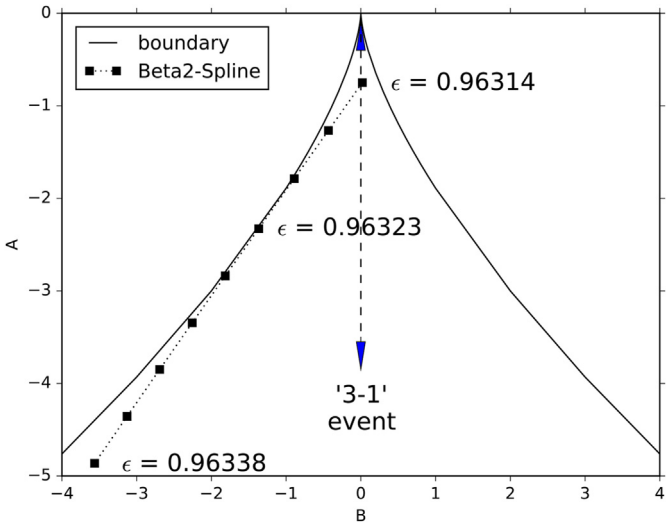


Fig. 8. Catastrophe theory representation of both a "3–1 termination" (dashed) and "2-2 crossover" (dotted) event. The curved boundary separates the region of three solutions from the single solution range.

We note, finally, that the same sub-manifold also exists in the ODF fit of a quartic Bézier to a hypoTrochoid curve [[12], Chap.8]. In this case we have four degrees of freedom instead of two, but otherwise the specification of $\{\mu_i\}$ and $\{\nu_i\}$ proceeds in the same way as above.

8. Modeling an avoided catastrophe

We wish to re-express these events in the language of catastrophe theory, which is visually more informative. In order to model events involving three critical points we define s, a one-dimensional streamline [21,5] joining the critical points, and assume that $F(s)$ is approximated by a quartic function along this path:

$$F = s^4/4 + As^2/2 + Bs. \tag{47}$$

In this equation all the parameters which affect only the scaling and positioning have been removed, leaving only two parameters which control the shape. This can be used to describe a *cusp catastrophe* [3,4]. The function F has three critical points when

$$B^2/4 + A^3/27 < 0. \tag{48}$$

Otherwise it has only one critical point. The boundary curve, Eq. (48), is shown in Fig. 8 and Fig. 1.

We wish to show that both the 3–1 termination and the 2-2 crossover events are represented by a tangential touching of the boundary curve in this figure, while a "normal" merge of two critical points is represented by a non-tangential crossing of this boundary. Consider a one-dimensional problem in which we search for critical points of $F(s)$ along the streamline s. Then the condition that **M** be singular at a critical point is expressed, Eq. (47), as:

$$d^2F/ds^2 = 3s^2 + A = 0. \tag{49}$$

The condition that $d\mathbf{a}/d\varepsilon$ be well-behaved, or the condition that $\mathbf{M}^+$ be simultaneously singular, is satisfied if

$$d^2F/dsd\varepsilon = sdA/d\varepsilon + dB/d\varepsilon = 0. \tag{50}$$

Combining them we obtain the constraint

$$dA/dB = -1/s = \pm\sqrt{-3/A} \tag{51}$$

which is just the expression for the slope of the catastrophe boundary.

As an application of this concept, consider the 3–1 termination event. In this case, two degenerate saddle points merge with a higher symmetry local minimum to produce a single saddle point, while $d\mathbf{a}/\,d\varepsilon$ is well-behaved. Since the two saddle points have the same rms error, we find that $B = 0$ in Eq. (47). Therefore the tangential touching of the boundary occurs at the origin. The additional constraint that is required in order to produce a tangential touching is provided by symmetry. The tangential touching event also coincides with a normal fold catastrophe at the origin to produce a single saddle point at $A > 0$. The path that the system follows is plotted as a vertical line in Fig. 8, which passes through the origin.

The 2-2 crossover event of Sec. 7 is more challenging. We wish to fit $F(s)$ to a sequence of ODF data obtained along the path joining three critical points in Fig. 7a, to see how the fit evolves as ε changes. For each ε we need at least five data points to do so. Three of these points are already known from the ODF optimization at the critical points, but two remain to be calculated. The problem is that these two points are no longer critical points; instead, they are points defined along an unknown streamline that joins a set of critical points [21]. The streamline is defined to be tangent to the gradient of F at all times. The uniqueness of the streamline was shown in Ref. [5]. However, calculating it exactly turns out to be difficult: in the vicinity of a critical point the typical behavior [[22], p.183] is that the streamline is a function of the type $p_1^{1/\lambda_1} \propto p_2^{1/\lambda_2}$, where $\mathbf{p}$ is a point in the transformed coordinate system defined by the eigenvectors of $\mathbf{M}$, λ_i are the eigenvalues, and the origin of $\mathbf{p}$ is at the critical point. As we approach a critical point these streamlines will exhibit very high curvature, particularly if the ratio $\lambda_2/\,\lambda_1$ is large, which will be the case if two critical points approach each other. However, we do not need the entire streamline, just a set of intermediate points between two critical points. We evaluate these approximately using the method of steepest descent [[23], p.121]: calculate a preliminary estimate of $\mathbf{a}$ by linearly interpolating between a saddle point and a local minimum. Then perform some iterations of the steepest descent method. Each iteration will increment $\mathbf{a}$ by the amount $\Delta\mathbf{a} = -\alpha\mathbf{L}$, where $\alpha = \mathbf{L}^t\mathbf{L}/\mathbf{L}^t\mathbf{M}\mathbf{L}$. Typically the point $\mathbf{a}$ will oscillate back and forth across the narrow trough that leads to the minimum [[23], Fig8.2]. We terminate the iteration when the gain α stabilizes, since any point along this streamline will do. Perform this for both pairs of minima/saddle points at each ε to obtain a total of five data points. Then fit a general quartic model of $F(s)$ and convert this to the standard form in Eq. (47). The resulting data are shown in Fig. 8. The figure shows the results of this method applied to the current Beta2-spline, where the steepest descent iteration was terminated after 20 iterations. The path length s was evaluated using the distance in the (d_1, d_2) plane. We see that there is a tangential touching event at $\varepsilon = 0.96323$, followed very closely by a fold catastrophe. Unfortunately, the particular numerical method we are using does not allow us to actually cross the boundary at the fold catastrophe, but it is clear that this crossing is imminent. A similar avoided catastrophe was obtained by applying this method to a clamped single-segment cubic Bézier fit (see Fig. 5.1 in Ref [[12]]) at $c \approx 3.59$, with the steepest descent iteration terminated after one iteration.

Fig. 8 confirms the supposition that a 2-2 crossing consists of a tangential touching of the catastrophe boundary. We also note that the sign of B has changed between the 2-2 crossing and the corresponding fold catastrophe. Therefore not the same pair of critical points is involved in these two events. The figure also clarifies the relationship between these events. Given the highly curved character of the catastrophe boundary, and given the fact that the relationship between A and B is nearly linear as ε varies, it is almost inconceivable that a 2-2 crossover event could occur without a corresponding fold catastrophe. This is an insight that we would not have obtained if we restricted ourselves to the use of only the determinants of $\mathbf{M}$ and $\mathbf{M}^+$.

9. Conclusion

When fitting a spline to a closed shape it is normal to see multiple solutions, which will appear as local minima and saddle points in the error function. If the shape of the object is changed by varying a single external paramater, then it may happen that these different types of solutions meet, leading to a topological change in which solutions either disappear or interchange their character. The outcome of these topological events can be determined by inspecting two different matrices which represent the second-order response of the curve-fitting error to changes in the curve-fitting parameters and the shape of the object. Events in which two solutions annihilate are quite common and do not require any special constraints to be met, other than the change in shape. Events in which two solutions survive and interchange character are rare, and require both of these response matrices to be simultaneously singular. In the context of ODF curve-fitting, a new algebraic constraint is defined, which allows this type of solution crossover to occur. It is also possible for solution bifurcations to occur as a result of symmetry-breaking, which is a relatively common occurrence. These three types of events: annihilation, crossover, and symmetry-breaking, can be analyzed algebraically using matrix determinants, but that does not lead to any understanding of the relationship between them. A more visual representation is obtained using catastrophe theory, which allows us to represent them as being either a simple crossing of a boundary (annihilation), a simultaneous crossing of two boundaries which have merged (symmetry-breaking), or a tangential touching of a boundary in which a catastrophe is avoided (crossovers). In the case of a crossover, the catastrophe theory representation makes it clear that a crossover is not only rare, but will almost certainly be closely associated with an annihilation event, so much so that it may be difficult to distinguish the two events.

Credit author statement

Alvin Penner: Conceptualization, implementation, writing.

Declaration of competing interest

I declare that I have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author is grateful to the reviewers and editors for their valuable comments, which were used to improve the quality of this paper.

The calculations were performed using the Java code at the repository: https://github.com/alvinpenner/Spiro2SVG/.

References

[1] Ahn SJ. Least squares orthogonal distance fitting of curves and surfaces in space, LNCS 3151. Berlin: Springer; 2004. https://doi.org/10.1007/b104017.
[2] Aubin J-P, Ekeland I. Applied nonlinear analysis. Mineola, N.Y.: Dover Books on Mathematics, Dover Publications; 2006.
[3] Stewart I. Catastrophe theory in physics. Rep Prog Phys 1982;45(2):185–221. https://doi.org/10.1088/0034-4885/45/2/002.
[4] Zeeman EC. Catastrophe theory, Scientific American 1976;234(4):65–83. https://doi.org/10.1038/scientificamerican0476-65.
[5] Antony C. Gradient flow line near birth-death critical points. J Topol Anal 2018;10:1–45. https://doi.org/10.1142/s1793525319500493.
[6] Gay DT. Functions on surfaces and constructions of manifolds in dimensions three, four and five. 2017. p. 1–20. https://arxiv.org/abs/1701.01711.
[7] Edelsbrunner H, Harer J, Mascarenhas A, Pascucci V, Snoeyink J. Time-varying reeb graphs for continuous space-time data. Comput Geom 2008;41(3):149–66. https://doi.org/10.1016/j.comgeo.2007.11.001.
[8] Kirby R. A calculus for framed links in s3. Invent Math 1978;45(1):35–56. https://doi.org/10.1007/bf01406222.
[9] Strogatz SH. Nonlinear dynamics and chaos. CRC Press; 2018. https://doi.org/10.1201/9780429492563.

[10] Golubitsky M, Schaeffer D. A theory for imperfect bifurcation via singularity theory. Commun Pure Appl Math 1979;32(1):21–98. https://doi.org/10.1002/cpa.3160320103.
[11] Bartels RH, Beatty JC, Barsky BA. An introduction to splines for use in computer graphics & geometric modeling. Los Altos: Morgan Kaufmann; 1987.
[12] Penner A. Fitting splines to a parametric function. Cham, Switzerland: Springer Briefs in Computer Science, Springer; 2019. https://doi.org/10.1007/978-3-030-12551-6.
[13] Ahn SJ, Rauh W, Cho HS, Warnecke H-J. Orthogonal distance fitting of implicit curves and surfaces. IEEE Trans Pattern Anal Mach Intell 2002;24(5):620–38. https://doi.org/10.1109/34.1000237.
[14] Bjorck A. Numerical methods for least squares problems. Philadelphia: SIAM; 1996. https://doi.org/10.1137/1.9781611971484.
[15] Penner A. Fitting a cubic bezier to a parametric function. Coll Math J 2019;50:185–96. https://doi.org/10.1080/07468342.2019.1583038.
[16] McWeeny R. Symmetry: an introduction to group theory and its applications. Mineola, N.Y.: Dover; 2002.
[17] de Boor C, Hollig K, Sabin M. High accuracy geometric hermite interpolation. Comput Aided Geomet Des 1987;4(4):269–78. https://doi.org/10.1016/0167-8396(87)90002-1.
[18] de Tisi F, Rossini M. Behavior of the beta-splines with values of the parameters beta2 negative. Comput Aided Geomet Des 1992;9(6):419–23. https://doi.org/10.1016/0167-8396(92)90040-V.
[19] Mascarenhas A, Snoeyink J. Isocontour based visualization of time-varying scalar fields. In: Mathematics and visualization. Springer Berlin Heidelberg; 2009. p. 41–68. https://doi.org/10.1007/b106657_3.
[20] Guillemin V, Pollack A. Differential topology. Englewood Cliffs: Prentice-Hall; 1974. https://doi.org/10.1090/chel/370.
[21] Gordon P. Paths connecting elementary critical points of dynamical systems. SIAM J Appl Math 1974;26(1):35–102. https://doi.org/10.1137/0126005.
[22] Lefschetz S. Differential equations: geometric theory. New York, N.Y.: Dover; 1977.
[23] Chong EKP, Zak SH. An introduction to optimization. Wiley; 2008. https://doi.org/10.1002/9781118033340.