

Rational cubic spirals

Donna A. Dietz^{a,*}, Bruce Piper^b, Elena Sebe^b

^aDepartment of Mathematics, Mansfield University, Mansfield, PA, 16933, USA

^bDepartment of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Received 22 September 2006; accepted 1 May 2007

Abstract

We consider the problem of finding parametric rational Bézier cubic spirals (planar curves of monotonic curvature) that interpolate end conditions consisting of positions, tangents and curvatures. Rational cubics give more design flexibility than polynomial cubics for creating spirals, making them suitable for many applications. The problem is formulated to enable the numerical robustness and efficiency of the solution-algorithm which is presented and analyzed.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Monotonic curvature; Rational Bézier cubics; Spiral; Polynomial

1. Introduction

In this paper, we study numerical methods that aid in the selection of rational cubics for applications where monotonic curvature is important. Since spirals are free of local curvature extrema, spiral design is an interesting mathematical problem with importance for both physical [8] and aesthetic applications [2]. Since rational Bézier cubics are common to all modern design systems [6] and offer more flexibility than polynomial Bézier cubics [13], it is convenient to describe rational cubic spirals so that spirals may be used in a variety of CAD systems. For other work on spirals with prescribed end conditions see [9] and [11], and the references therein. Recently, in [5] and [4], numerical techniques were used to study parametric Bézier cubic spirals. In this paper, that work is continued by studying parametric *rational* Bézier cubic spirals, and a fast, robust algorithm is presented for finding these so that they interpolate given end conditions. (For the remainder of this article, a *cubic* is a parametric polynomial planar cubic, and a *rational cubic* is a parametric rational planar cubic.)

First, Section 2 gives background and notation used in this paper. Then, Section 3 formulates the problem of optimizing rational cubic spirals by choosing a useful expression for the

free parameters. Section 4 describes the algorithm used for finding the optimal rational cubic spiral. Section 5 contains comments on the numerical results of this investigation.

2. Background and notation on spirals and rational cubics

2.1. Rational cubic curves

If a *rational cubic* spiral is to exist satisfying given tangential and curvature end conditions, necessarily, some *rational cubic* must exist which satisfies those end conditions. (Naturally, the question would still remain as to whether or not it is a *spiral*!)

Rational cubic curves are represented as

$$\mathbf{f}(t) = \frac{\sum_{v=0}^3 w_v \mathbf{b}_v B_v^3(t)}{\sum_{v=0}^3 w_v B_v^3(t)}, \quad 0 \leq t \leq 1,$$

$$B_i^n(t) = \binom{n}{i} (1-t)^{n-i} t^i, \quad (1)$$

where \mathbf{b}_v are the four planar control points, and w_v are the four scalar weights. (If all the weights are set equal, the resulting curve is a cubic.) A complete discussion of rational Bézier cubics may be found in [6] and [7]. Since it does not alter curvature properties (except by a constant scale), throughout this paper it is assumed that $\mathbf{b}_0 = (0, 0)$, and $\mathbf{b}_3 = (1, 0)$.

* Corresponding author. Tel.: +1 570 662 4705; fax: +1 570 662 4137.

E-mail addresses: ddietz@mansfield.edu (D.A. Dietz), piperb@rpi.edu (B. Piper), sebee@rpi.edu (E. Sebe).

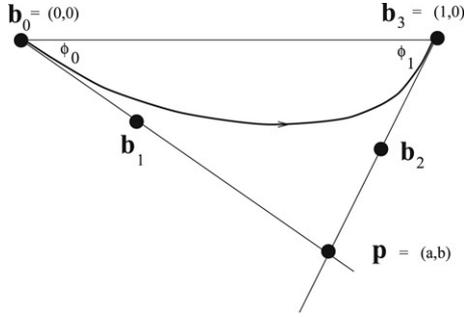


Fig. 1. Tangential conditions are specified with ϕ_0 and ϕ_1 .

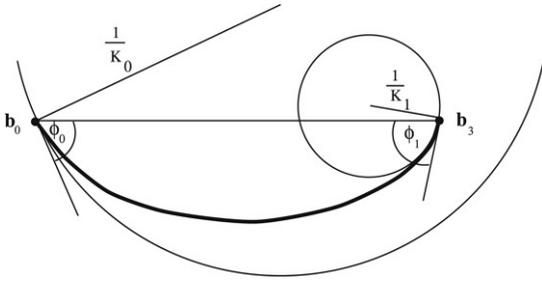


Fig. 2. Circles of curvature at \mathbf{b}_0 and \mathbf{b}_3 .

This would seem to leave eight degrees of freedom for design, corresponding to the four weights and two remaining control points in the plane, \mathbf{b}_1 and \mathbf{b}_2 (which are both assumed to lie in the fourth quadrant so as to allow for non-negative curvature). However, as explained in [7], two degrees of freedom are lost, as representation of rational cubics is not unique. Any two weights may be arbitrarily chosen (not equal to zero) with no loss of freedom. Therefore, six degrees of freedom remain for design purposes. To ensure unique representation, the conditions $w_1 = 2/3$ and $w_2 = 2/3$ are imposed as was done in [1].

As shown in Fig. 1, the tangent lines for the rational cubic at $t = 0$ and $t = 1$ and the horizontal axes form a triangle with angles ϕ_0 at the $t = 0$ corner and ϕ_1 at the $t = 1$ corner which are the tangential end conditions for the rational cubic. Curves whose tangent vectors turn through smaller angles are more likely to be useful in many applications, so the assumption is made that $0 < \phi_0 < \pi/2$, and $0 < \phi_1 < \pi/2$. Thus, the intersection of the tangent lines shown in the Fig. 1 exists and is given by

$$\mathbf{p} = (a, b), \quad \text{where } a = \frac{\cos \phi_0 \sin \phi_1}{\sin(\phi_0 + \phi_1)},$$

$$\text{and } b = -\frac{\sin(\phi_0) \sin(\phi_1)}{\sin(\phi_0 + \phi_1)}. \quad (2)$$

The lengths of the lower two sides of the triangle in Fig. 1 are $d_0 = \sin(\phi_1)/\sin(\phi_0 + \phi_1)$ (for the side touching the origin), and $d_1 = \sin(\phi_0)/\sin(\phi_0 + \phi_1)$ for the side touching $(1, 0)$. The ratios

$$f_0 = \frac{|\mathbf{b}_1 - \mathbf{b}_0|}{d_0} \quad \text{and} \quad f_1 = \frac{|\mathbf{b}_2 - \mathbf{b}_3|}{d_1} \quad (3)$$

are used extensively in this paper so that the four variables ϕ_0, ϕ_1, f_0 , and f_1 or the four variables a, b, f_0 , and f_1 may represent the same four degrees of freedom as \mathbf{b}_1 and \mathbf{b}_2 . Thus, \mathbf{b}_1 and \mathbf{b}_2 are replaced by

$$\mathbf{b}_1 = f_0 \mathbf{p} \quad \text{and} \quad \mathbf{b}_2 = (1 - f_1) \mathbf{b}_3 + f_1 \mathbf{p}. \quad (4)$$

Both f_0 and f_1 are constrained to be between zero and one. This (with the angle restrictions imposed above) implies that the rational cubic is convex and hence free from inflections.

In terms of a, b, f_0, f_1, w_0 and w_3 , the rational cubic is given by

$$x(t) = \frac{w_3 t^3 + 2(1 - f_1(1 - a))(1 - t)t^2 + 2f_0 a(1 - t)^2 t}{w_3 t^3 + 2(1 - t)t^2 + 2(1 - t)^2 t + w_0(1 - t)^3}, \quad (5)$$

and

$$y(t) = \frac{2f_1 b(1 - t)t^2 + 2f_0 b(1 - t)^2 t}{w_3 t^3 + 2(1 - t)t^2 + 2(1 - t)^2 t + w_0(1 - t)^3}. \quad (6)$$

2.2. Spirals

If a rational cubic *spiral* is to exist satisfying given tangential and curvature end conditions, necessarily, some *spiral* must exist which satisfies those end conditions. (Naturally, the question would still remain as to whether or not it is a *rational cubic*!) The curvature end conditions are called K_0 and K_1 just as the tangential end conditions are called ϕ_0 and ϕ_1 . If $K_0 \neq 0$, the circle passing through \mathbf{b}_0 with radius $1/K_0$ that is tangent to the line $\mathbf{b}_0\mathbf{b}_1$ and lies on the same side of this line as does \mathbf{b}_3 will be called the circle of curvature at \mathbf{b}_0 , and analogously for \mathbf{b}_3 .

For the purposes of this work, spirals are defined to be planar arcs having both non-negative curvature and continuous non-zero derivative of curvature. Thus, spirals have monotonic curvature and are free from inflections. The formula for curvature of a parametric curve is

$$K(t) = \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}, \quad (7)$$

where x and y are functions of t . This formula shows the non-linearity of the task of using rational cubics and ensuring monotonicity of curvature.

Without loss of generality, the spirals studied will be of increasing curvature for increasing values of t . From Theorem 3.18 in [10], there exists some convex spiral arc of increasing curvature interpolating the end conditions $\mathbf{b}_0 = (0, 0), \phi_0, K_0$ and $\mathbf{b}_3 = (1, 0), \phi_1, K_1$ with $\phi_0 \in (0, \pi/2)$, and $\phi_1 \in (0, \pi/2)$ with $0 < K_0 < K_1$ if and only if the circle of curvature at $(0, 0)$ contains the circle of curvature at $(1, 0)$, and $0 < \phi_0 < \phi_1 < \pi/2$. (These circles of curvature are shown in Fig. 2.) The latter condition is equivalent to the constraints that $0.5 < a < 1$, and $b < 0$ in Fig. 1. Using elementary geometry, the former condition is formulated by imposing a first constraint that is derived from the tangential contact of the two circles of curvature and a second constraint ensures that the circle of

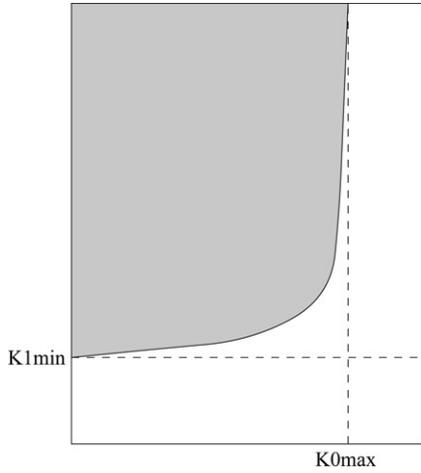


Fig. 3. Region in (K_0, K_1) space (for fixed ϕ_0 and ϕ_1) where each $(\phi_0, \phi_1, K_0, K_1)$ quadruple can be interpolated by spirals with $K_{0 \max} = 2 \sin \phi_0$ and $K_{1 \min} = (1 - \cos(\phi_0 + \phi_1)) / \sin \phi_0$.

curvature at \mathbf{b}_0 actually contains \mathbf{b}_3 . These two constraints lead to the inequalities, $K_0 < 2 \sin(\phi_0)$ and

$$K_1 > \frac{2(1 - \cos(\phi_0 + \phi_1)) - 2K_0 \sin(\phi_1)}{2 \sin(\phi_0) - K_0}. \quad (8)$$

For a spiral, when $K_0 = 0$ the tangent line at $(0, 0)$ must support the circle of curvature at $(1, 0)$ and this again leads to Inequality (8). For fixed ϕ_0 and ϕ_1 , Inequality (8) describes a region in (K_0, K_1) space above a hyperbola where each $(\phi_0, \phi_1, K_0, K_1)$ quadruple can be interpolated by some spiral, as shown in Fig. 3.

Since there is no closed form for the roots of the derivative of the curvature for a rational cubic, a numerical solution is pursued.

3. Formulating the problem of optimizing a rational cubic spiral by choosing useful expressions of free parameters

3.1. Parameter selection

In [5] and [4], the problem of designing cubic spirals was approached by calculating numerous cases numerically. The tangential end conditions were set, and curvature end conditions which yielded spirals were extracted and tabulated.

For rational cubics, the situation is more complicated due to the number of parameters, so analysis begins with the equation for curvature. The problem is how best to choose the parameters involved. As explained in Section 2.1, if a and b are prescribed, there are four remaining degrees of freedom, $f_0, f_1, w_0,$ and w_3 with which to control K_0 and K_1 .

The relations between these variables are derived from the formulas for curvature and are given by

$$K_0 = \frac{-w_0 b(1 - f_1)}{f_0^2 (a^2 + b^2)^{3/2}}, \quad (9)$$

and

$$K_1 = \frac{-w_3 b(1 - f_0)}{f_1^2 ((1 - a)^2 + b^2)^{3/2}}. \quad (10)$$

If tangential and curvature end conditions are prescribed, $a, b, K_0,$ and K_1 are known, so for each (a, b, K_0, K_1) quadruple, values of $f_0, f_1, w_0,$ and w_3 are sought so that the above equations are satisfied, and the resulting rational cubic is a spiral. Since Eqs. (9) and (10) can be solved for any two of their variables in terms of the others, there are two degrees of freedom remaining after the curvature constraints are imposed. The task is thus reduced to selecting the pair of the remaining two degrees of freedom to produce a spiral if possible. (This is faster than searching over a 4-D space while checking the constraints.)

It is advantageous to solve Eqs. (9) and (10) for f_1 and w_3 with independent variables f_0 and w_0 . This gives

$$f_1 = \frac{w_0 b + K_0 f_0^2 (a^2 + b^2)^{3/2}}{w_0 b}, \quad (11)$$

and

$$w_3 = \frac{-K_1 f_1^2 ((a - 1)^2 + b^2)^{3/2}}{(1 - f_0) b}, \quad (12)$$

where the expression for f_1 may be substituted into the expression for w_3 directly, or, when this is done numerically, Eq. (11) may be simply calculated before Eq. (12).

At first glance, it may seem that it would be easier to simply leave two of the weights (either w_0 and w_3 or w_1 and w_2) as the degrees of freedom. However, after some trial and error, we found that the above form leads to greater flexibility, simpler formulations and numerical stability. The reasons it is better to leave f_0 and w_0 (rather than two weights) as the last two free variables are as follows:

- The formulas are quite simple.
- Since the curvature will be increasing and positive, there is no need to maintain symmetry between the use of f_0, w_0 and f_1, w_3 .
- When the curvature K_0 is set to zero in Eq. (9), the consequence should be that $f_1 = 1$ which is exactly what happens in Eq. (11). However, some of the other formulations force w_0 to become zero, and this is not useful from a numerical standpoint or a design standpoint.
- Since the curvature will be increasing and positive, the curvature at $t = 1$ is never zero and hence there is no need to make $f_0 = 1$. So the denominator in Eq. (12) will not be zero for $b < 0$. A small denominator in this expression would imply that w_3 is large, but since it is important for the numerical stability of the curve representation itself to keep w_3 bounded, this restriction on w_3 imposes a bound on the denominator of Eq. (12).
- In a similar manner, the condition that the denominator in Eq. (11) not be small is ensured by making certain that neither $|b|$ nor w_0 are small. This, again, is consistent with numerical stability of the curve itself.

The constraint that $0 \leq f_1 \leq 1$ leads to a lower bound on w_0 which is given by

$$w_{0 \min} = \frac{K_0 f_0^2 (a^2 + b^2)^{3/2}}{-b}. \quad (13)$$

Table 1
Simple cases for finding M_{exact} (without the $f_0 f_1 (1 - f_0)$ factor) for certain piecewise linear curvatures

	$K(0)$	$K(0.5)$	$K(1)$	Total Variation	$K_1 - K_0$	M_{exact}
A spiral	0	0.5	0.6	0.6	0.6	0.2
Boundary case	0	0.5	0.5	0.5	0.5	0
Not a spiral	0	0.5	0.4	0.6	0.4	-0.2

There is no known theoretical upper bound on w_0 , but in practice, for the majority of cases we have computed, if there are rational cubic spirals for a given $(\phi_0, \phi_1, K_0, K_1)$ quadruple, there will also be some with w_0 less than 6 (usually closer to 1 or 2). In no cases did we observe w_0 over 9, but rare instances may exist. More specifically, the “optimal” rational cubic spiral will have a low value in this range for its w_0 . (The measure for “optimal” is discussed in Section 3.2.) Thus, a reasonable bounded region of (f_0, w_0) space can be formed to numerically search for a rational cubic spiral.

3.2. A measure of the quality of the spiral

To select f_0 and w_0 , a measure, M_{exact} , is defined to describe how close a curve is to being a spiral. A numerical optimization is done on M_{exact} , so the measure must have its largest positive values for numerically stable spirals. From an artistic standpoint, there are various attributes that a designer might wish to emphasize (such as maximum rotational symmetry or minimizing total variation in some higher order derivative of K — with the cubic being parameterized however desired, possibly by arclength) and any of these goals would be compatible with the methods described here. Those attributes could be worked into a new measure, so long as the desired attributes did not cause numerical instabilities. For the measure used in this work, spirals with larger minimum K' and smaller maximum K' were preferred by the algorithm. This is discussed in Section 5.2.4.

The measure is a function mapping rational Bézier curves to the real numbers and is defined so that negative values result when the curve is not a spiral and positive values result when the curve is a spiral. It works well to use different functions depending upon whether or not the curve is a spiral, because this improves how the measure drives the optimization algorithm. Again, since the measure is maximized numerically, it should have its largest positive values for numerically stable spirals.

Given a rational cubic, the first step in computing the measure is to compute the curvature, $K(t)$, from Eq. (7), over $0 \leq t \leq 1$ as in the previous section. Then, the measure M_{exact} is given by

$$\begin{aligned}
 M_{\text{exact}}(a, b, K_0, K_1, f_0, w_0) &= M_{\text{exact}}(f_0, w_0) \\
 &= \begin{cases} f_0 f_1 (1 - f_0) \min(K'(t)) & \text{if } K'(t) > 0 \\ & \text{for all } 0 \leq t \leq 1 \\ K_1 - K_0 - \int_0^1 |K'(t)| dt & \text{otherwise.} \end{cases} \quad (14)
 \end{aligned}$$

Observe that the measure is positive in the first case and the factors f_0 , f_1 and $(1 - f_0)$ are included. Numerical experience

has shown that these are useful in driving the algorithm to choose a reasonable spiral from a numerical standpoint.

The measure is non-positive in the second case, where it measures the discrepancy between the total variation of the curvature and the minimum possible value of the total variation, K_1 minus K_0 . The neutral case occurs when $K'(t) \geq 0$ and $K'(t) = 0$ for at least one t in $[0, 1]$, hence $\min(K'(t)) = 0 = K_1 - K_0 - \int_0^1 |K'(t)| dt$. The measure $M_{\text{exact}}(f_0, w_0)$ is to be maximized for each (a, b, K_0, K_1) quadruple. Since M_{exact} cannot be directly computed, it must be approximated, as described in Section 4.2.

3.3. Short demonstrations for measure M_{exact}

In Table 1, three simple examples are shown to demonstrate the computation of the measure M_{exact} in the event that the curvature is piecewise linear with two linear segments joining at $t = 1/2$. These examples are shown just to illustrate the measure as the curvature for rational cubics is obviously not piecewise linear. Further, we do not include the factor of $f_0 f_1 (1 - f_0)$.

For each example, the curvature is found at $t = 0$, $t = 0.5$, and $t = 1$, so the step-size is 0.5. The total variation is found based on those partition points, and $K_1 - K_0$ is also found. In the first case, the total variation is equal to $K_1 - K_0$, and the measure M is the smallest $\Delta K / \Delta t$ which, of course, is positive. In the last case, the total variation exceeds $K_1 - K_0$, and the total variation is subtracted from $K_1 - K_0$, giving a negative value for M_{exact} . In the middle case, where M_{exact} is zero, the calculation for both the above cases are zero and this occurs precisely when there are consecutive partition points with the same curvature value but no consecutive partition points indicate a decreasing curvature.

4. An algorithm for finding optimal rational cubic spirals

This section describes the algorithm for finding optimal rational cubic spirals. As before, the endpoints are fixed at $\mathbf{b}_0 = (0, 0)$ and $\mathbf{b}_3 = (1, 0)$, and $w_1 = w_2 = 2/3$. The algorithm takes as input the values of ϕ_0 , ϕ_1 , K_0 , and K_1 as well as a positive integer N (used to set a mesh size) which will be used in creating M , an approximation to the measure M_{exact} . The output gives f_0 and w_0 where M is found to be greatest. The values of f_1 and w_3 are computed from f_0 and w_0 by Eqs. (11) and (12). This is sufficient information to construct the rational Bézier using Eqs. (1), (4) and (2).

The algorithm has three subroutines, one to calculate curvature, one for determining M , and one for optimizing M .

4.1. Calculating the curvature

For rational cubics, the calculation of curvature is both more numerically stable and more efficient than the computation of its derivative. Thus, for values of t other than zero and one, the approximation of the measure in the next subsection is based mostly on curvature values.

The subroutine for calculating the curvature takes as inputs f_0, w_0, K_0, K_1, a and b as well as t . These uniquely define a rational cubic using Eqs. (5), (6), (11) and (12) and provide the point at which the curvature is to be calculated. The routine returns the curvature at t using Eq. (7). The calculations for the first and second derivatives are optimized with automatically generated code produced by Maple™ [12] from the symbolic derivatives.

Due to possible extrema hiding between mesh points (discussed in Section 4.2, the derivatives of curvature at the two endpoints ($t = 0$ and $t = 1$) are found. While the evaluation of the derivative of the curvature for arbitrary values of t is computationally unstable and time-consuming, at $t = 1$ and $t = 0$ it is simple, making these two additional computations reasonably fast. This subroutine finds $K'(0)$ and $K'(1)$ by combining formulas for the first, second and third derivatives of the rational cubic spiral at $t = 0$ and $t = 1$ together with the derivative of the formula given in Eq. (7).

4.2. Calculating M and optimizing spirals

This subroutine approximates the measure M_{exact} defined in Eq. (14) using the inputs f_0, w_0, K_0, K_1, a, b and a sample size N . We discuss two approximations, a simple approximation called \bar{M} and then an improved approximation called M . The approximations are based on a partition of $[0, 1]$ given by

$$0 = t_0 < t_1 < t_2, \dots < t_n = 1$$

where n will be based on the input N and will be discussed further below. The curvature of the rational cubic obtained from $f_0, w_0, K_0, K_1, a,$ and b is calculated at a set of t values in $[0, 1]$ to obtain a list of values, $\kappa_i = K(t_i)$ where $i = 0, \dots, n$. Once the κ_i values are computed, the routine computes $\Delta\kappa_i = \kappa_i - \kappa_{i-1}$ and $\Delta t_i = t_i - t_{i-1}$ for $i = 1, \dots, n$. The measure is then approximated by checking to see if all the $\Delta\kappa_i > 0$ in which case

$$\bar{M}(f_0, w_0) = f_0 f_1 (1 - f_0) \min_{1 \leq i \leq n} \left(\frac{\Delta\kappa_i}{\Delta t_i} \right) \quad (15)$$

which approximates the minimum slope. If at least one $\Delta\kappa_i$ is negative, the routine computes an approximation of the second case in Eq. (14). This is done numerically using a Riemann sum for the integral, where the derivative of the curvature to K is approximated by $\frac{\Delta\kappa_i}{\Delta t_i}$.

$$\begin{aligned} \bar{M}(f_0, w_0) &= K_1 - K_0 - \sum_{i=1}^n \left| \frac{\Delta\kappa_i}{\Delta t_i} \right| \Delta t_i \\ &= K_1 - K_0 - \sum_{i=1}^n |\Delta\kappa_i|. \end{aligned} \quad (16)$$

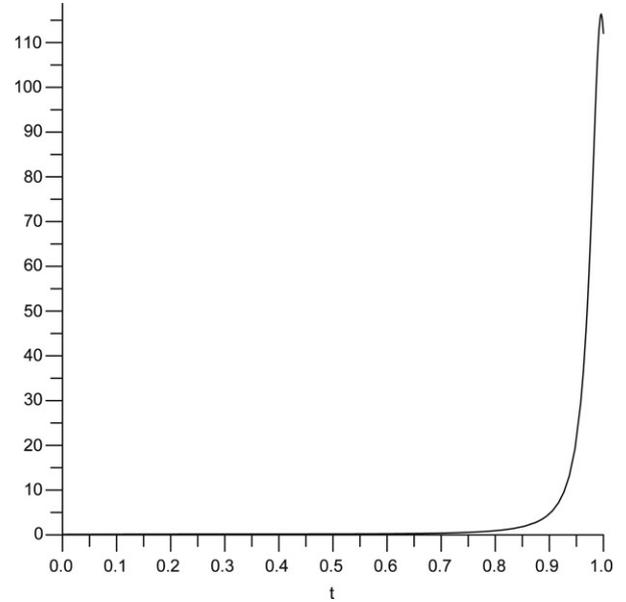


Fig. 4. A spike in the curvature which can hide between mesh points.

The obvious way to choose the partition of $[0, 1]$ is to do so uniformly and set $t_i = i/N$ (and $n = N$). However, experience shows that regardless of how large N is chosen, local maxima and minima can hide between the mesh points. Luckily, this usually occurs only between the last two mesh points as a local maximum, causing a very tall “spike” to appear in the graph of the curvature (as in Figs. 4 and 5). (Less common are local minima occurring between the first two mesh points.) This spiking phenomenon is discussed further in Section 5.

We discuss two methods by which the approximated measure can be designed to account for the spikes. The first method to deal with the spikes is to include extra values near 0 and 1 in the list of t values at which the curvature is computed. Thus the interval $[0, 1]$ would first be partitioned into N equal subintervals and j additional t values would be included in the first and last subintervals resulting in a partition of size $n = N + j$. If this method is used in isolation, it is best to cluster these added points close to 0 and 1.

The second method to deal with the spikes is to consider the derivatives of curvature at $t = 0$ and $t = 1$ as part of the measure. In this case, the t values are chosen by creating a uniform partition of $[0, 1]$ of size $1/N$ and then to also compute the derivatives of the curvature at $t = 0$ and $t = 1$ ($K'(0)$ and $K'(1)$.) The measure is then approximated by checking to see if $\Delta\kappa_i > 0$, for $i = 1 \dots N$ and if $K'(0) > 0$ and $K'(1) > 0$ in which case

$$\begin{aligned} M(f_0, w_0) &= f_0 f_1 (1 - f_0) \\ &\times \min \left(K'(0), \frac{\Delta\kappa_1}{\Delta t_1}, \frac{\Delta\kappa_2}{\Delta t_2}, \dots, \frac{\Delta\kappa_N}{\Delta t_N}, K'(1) \right) \end{aligned} \quad (17)$$

which approximates the minimum slope. If at least one $\Delta\kappa_i$ is negative or if $K'(1) < 0$ or $K'(0) < 0$, the routine computes

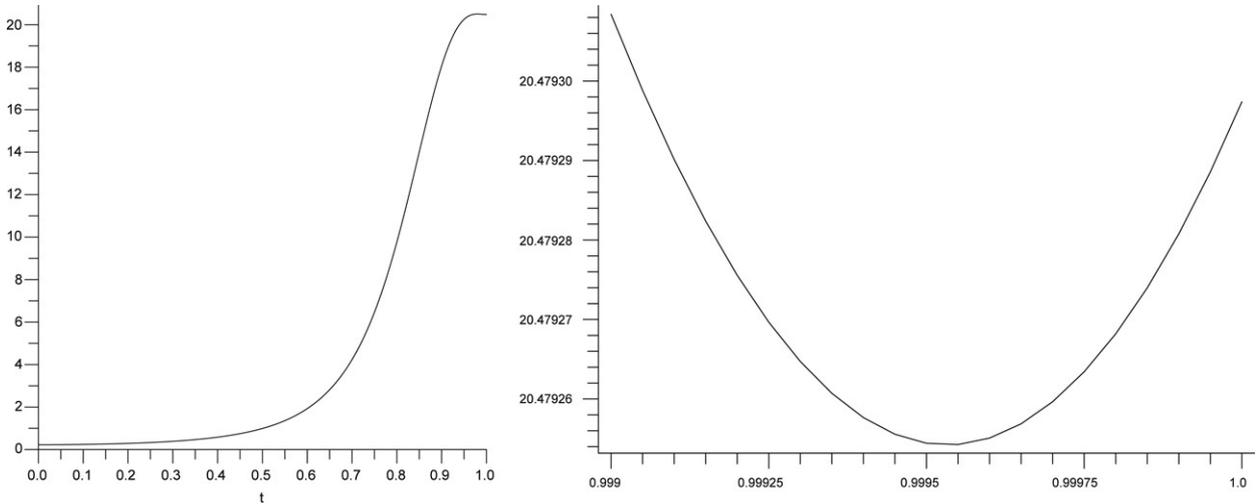


Fig. 5. A rare local minimum effect is hidden near $t = 1$.

$$M(f_0, w_0) = K_1 - K_0 - \left(\sum_{i=1}^N |\Delta\kappa_i| \right) + \min(0, K'(0)) + \min(0, K'(1)) \quad (18)$$

which is an approximation of the second case in Eq. (14) with a negative penalty if $K'(0) < 0$ or $K'(1) < 0$.

In practice, we use the measure M along with N additional points in each of the first and last subintervals of a uniform partition of size N , typically with $N = 15$. The use of $K'(0)$ and $K'(1)$ does well in assuring that the curve that gives rise to the spike in Fig. 4 is not labeled as a spiral, while the added points near $t = 1$ and $t = 0$ work better for the double spike in Fig. 5.

4.3. Optimizing M

As inputs, this subroutine takes values for a, b, K_0 , and K_1 and a starting point, (f_0^0, w_0^0) in the search space along with a grid size m and a tolerance ϵ . The outputs are f_0 and w_0 values where the measure is found to be greatest.

In the routine, an adaptive optimization procedure is used which begins at a starting value, (f_0^0, w_0^0) , searching an $m \times m$ grid, with a total starting width of one and height of $2w_0^0$. The value of M is computed for each point in the grid. Then, at each iteration, the $m \times m$ grid is recentered on the new point where the optimal measure occurs, and the width and height are halved. The routine continues until the distances between the points in the $m \times m$ grid are less than a preset tolerance ϵ for both f_0 and w_0 .

Thus, with the refinement allowing the search space to move upward, the total region searched is $0 \leq f_0 \leq 1$ and up to $0 \leq w_0 \leq 3w_0^0$. As discussed in Section 3.1, no spirals that optimize our measure have been found for $w_0 \geq 9$, so we typically choose $w_0^0 = 3$. We chose to start at lower values of w_0 as that is where the majority of spirals are found.

For each (f_0, w_0) point under consideration, the inequality $w_0 > w_{0 \min}$ where $w_{0 \min}$ is given by Eq. (13) is checked. If the inequality fails, the point is rejected from future consideration for the optimal measure. This method works

better than basing the lower end of the search region on the formula for $w_{0 \min}$ and using a non-rectangular grid because $w_{0 \min}$ can get large for $|b|$ small and f_0 near 1 which distorts the search grid and produces unfavorable results.

To further insure that a spiral is produced, when the optimization routine completes, the measure routine is run for the final curve but this time with a larger input value of N (usually 200). This includes many more curvature samples and occasionally rejects curves that would otherwise be labeled as spirals. We found that it is usual that in these cases, no rational cubic spiral exists. So greater speed is achieved by using a relatively small value of N (around 15) for the execution of the algorithm and then a larger value of N only at the completion of the algorithm for a final check.

For the bulk of the test cases we ran, we used a value of $m = 12$, set the tolerance $\epsilon = 0.01$, $f_0^0 = 0.5$, and as mentioned above, $w_0^0 = 3$. This resulted in an algorithm that runs in under a hundredth of a second for one case and is robust enough to handle the vast majority of cases. Thus, a more sophisticated optimization routine is not needed.

5. Comments on numerical results

5.1. Discretization and stability issues

We now discuss further the spiking phenomenon we described in Section 4.2. In Fig. 4, a sequence of curvatures for values of t chosen uniformly in the interval $[0, 1]$ will be increasing for even a fairly fine uniform mesh. Naturally, the interval may be increasingly subdivided, but there may still be spikes in the curvature which may be aliased and hence not be visible. In fact, the search algorithm seems to encourage this to happen. This happens almost entirely near $t = 1$ in the interval. Occasionally, a slight curvature minimum occurs near $t = 0$.

Two conditions often lead to this spiking phenomenon: a small value for ϕ_0 used in conjunction with a large value for ϕ_1 , and a small K_0 value used in conjunction with a relatively large K_1 value. In Fig. 4, both of these conditions hold, leading

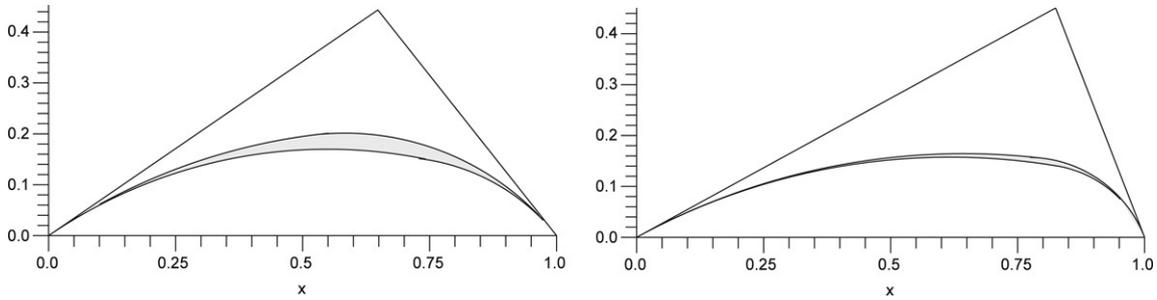


Fig. 6. Two examples of crescent-like regions containing all spirals satisfying tangential and curvature constraints.

to this spiking phenomenon. It follows from the constraint that $K_0 < 2 \sin(\phi_0)$ and Inequality (8) that only small K_0 and large K_1 pairs are potential spirals. (This is because since ϕ_0 is small the denominator of the right hand side of Inequality (8) is very small, but the numerator is not near zero because K_0 is also small and the $1 - \cos(\phi_0 + \phi_1)$ term dominates for large ϕ_1 .) It is easy to visualize why this is always the case for small ϕ_0 and large ϕ_1 . The rational cubic begins nearly flat and is confined to a thin triangle, but it must “turn” suddenly in order to match the tangent condition at $t = 1$. Fig. 6 shows a typical constraining region for two sets of ϕ_0, ϕ_1, K_0, K_1 sets.

The bottom boundary of the crescent-shaped region in Fig. 6 is a bi-arc. The left arc is from the required circle of curvature (with radius $\frac{1}{K_0}$) at $(0,0)$. Its right arc is from a circle satisfying the following two conditions. First, its tangent at $(1,0)$ makes an angle of ϕ_1 with the horizontal axis, and, second, it must touch tangentially that aforementioned circle of curvature originating at $(0,0)$. The upper boundary is also a bi-arc. Its right arc is from the required circle of curvature (with radius $\frac{1}{K_1}$) at $(1,0)$. The left arc is from a circle satisfying the conditions that its tangent at $(0,0)$ makes an angle of ϕ_0 with the horizontal axis, and that it touches tangentially the aforementioned circle of curvature originating at $(1,0)$.

A much less common phenomenon is to have both a local maximum and a local minimum close to $t = 1$. This is one way a non-monotonic curvature can hide numerically, even with the testing of the derivative of the curvature at $t = 1$. Fig. 5 demonstrates this. Note that for this particular example, the curve itself would be very close to being a spiral, as reflected by measure M having an extremely small magnitude. This example indicates the need for more points near $t = 1$ along with the derivative of curvature.

To see the impact of the value of N on the algorithm, we used the measure M as indicated in Section 4.2 and tried 156 000 test cases. With a value of $N = 15$, the algorithm found 61 293 spirals, whereas with $N = 20$, the algorithm found 61 931 spirals, for a very slight increase of around 1%.

We found the measure M outperformed all other approximations we tried. For example, we considered higher order integration techniques (e.g. Simpson’s rule) for approximating the integral in M_{exact} . Since the vast majority of these techniques have positive weights, they can do no better in detecting non-monotonic curvature than the measure M . As discussed in Section 4.1, the exact derivative of curvature is not computed but instead we use approximations to the derivative. Thus the

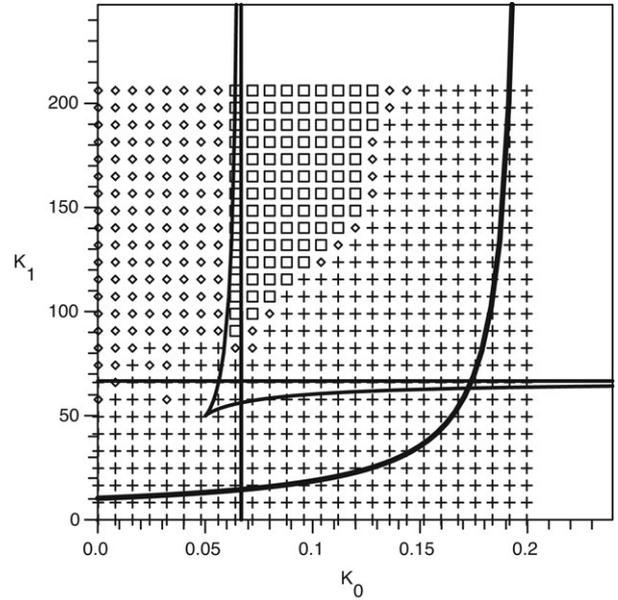


Fig. 7. Diagnostic plots in curvature space for $\phi_0 = 0.1$ and $\phi_1 = 1.5$.

higher order integration techniques do not actually result in a higher order approximation of the integral and in our experiments performed no better than the Riemann sum. We also experimented with an adaptive integration technique to approximate M_{exact} that tried to add more points to interior subintervals. This also worked no better than the approximation M . This is probably due to the fact that, as discussed in Section 4.2, the adaptation to the end intervals has already been made in computing M .

The measure M generally behaves well with respect to small changes in f_0 and w_0 . However, when the magnitude of M is large and M is negative, radical changes in M can occur for tiny perturbations in f_0 and w_0 . But the cases of greatest interest are when the magnitude of M is small and for those cases, M behaves quite reasonably. For example, numerical results indicate that when $-0.1 < M < 0$, a perturbation of ± 0.001 in f_0 and/or w_0 never produces a change of greater than 1 in M , and furthermore only results in a change in M greater than 0.1 in 176 of 498 101 sample cases. When the larger changes in M do occur they are usually not near the optimal value in the search space, making the algorithm quite well behaved in the vast majority of cases.

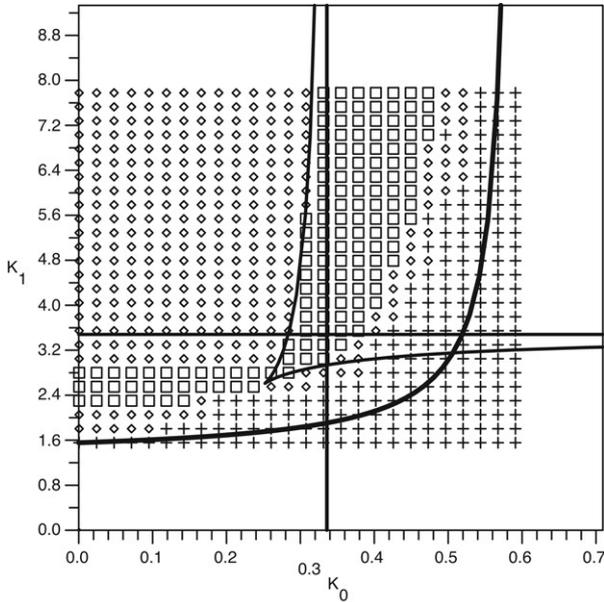


Fig. 8. Diagnostic plots in curvature space for $\phi_0 = 0.3$ and $\phi_1 = 0.7$.

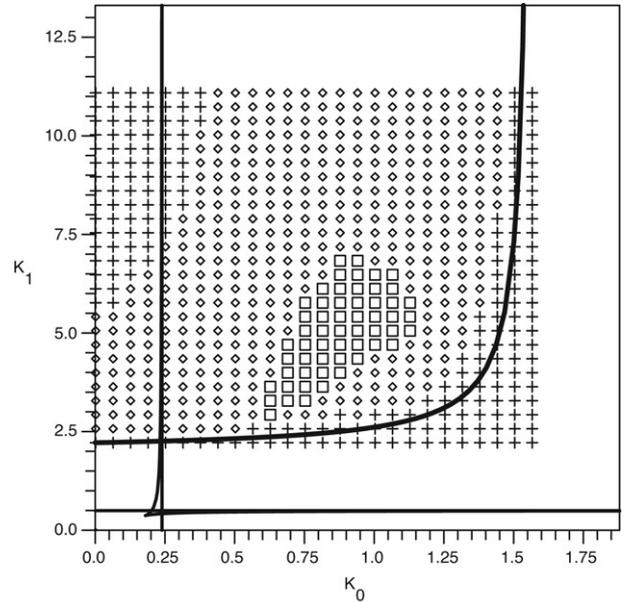


Fig. 10. Diagnostic plots in curvature space for $\phi_0 = 0.9$ and $\phi_1 = 1.4$.

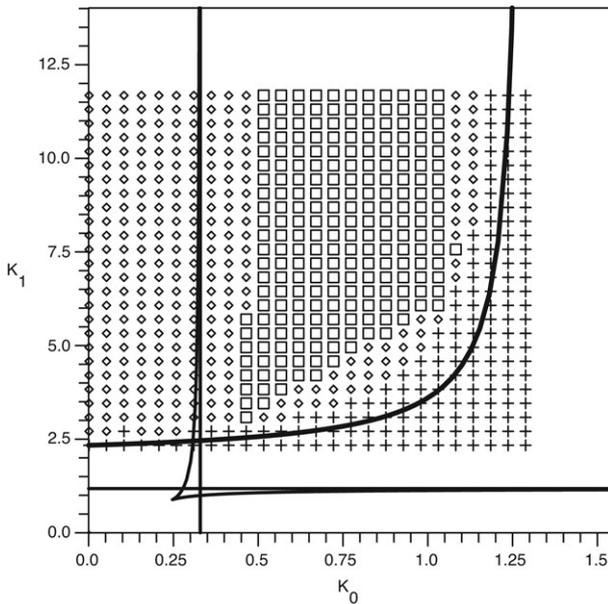


Fig. 9. Diagnostic plots in curvature space for $\phi_0 = 0.7$ and $\phi_1 = 1.4$.

5.2. Cubic diagnostic tools aiding rational cubic analysis

It is of interest to compare the rational cubic spiral findings to cubic spiral findings.

Figs. 7–10 are explained over the next few subsections. In these figures, cubic spirals and rational cubic spirals are compared by illustrating the (K_0, K_1) coordinates that corresponds to the spiral. The analysis of cubic spirals is fully described in [4]. Because every cubic spiral is a rational cubic spiral, the set of (K_0, K_1) values where a rational cubic spiral is found should be a superset of those where a cubic spiral exists. Furthermore, the cubic spiral analysis tools shed light on why the rational cubic spiral solutions may be exhibiting certain attributes.

5.2.1. Numerical comparison of cubics with rational cubics

For the over 156 000 cases, the algorithm (with $N = 15$) found 61 293 rational cubic spirals, and in 40 976 of these cases, there exists no corresponding cubic spiral. This is a reasonable measure of the greater flexibility of rational cubic spirals over cubic spirals. Some of the test points are shown as squares, diamonds, and crosses in Figs. 7–10.

The points in these four figures are either crosses (indicating that neither a cubic spiral nor a rational cubic spiral could be found for the (K_0, K_1) pair) or diamonds (indicating that the algorithm found rational cubic spirals but no cubic spiral exists), or squares (indicating the presence of a cubic spiral, which is, of course, also a rational cubic spiral, and usually many additional rational cubic spirals). For all but 8 cases of the over 156 000 of ϕ_0, ϕ_1, K_0, K_1 test values, the algorithm for the rational cubic spirals succeeded in producing a spiral when a cubic spiral existed. The exceptions are due to the fact that the algorithm for finding cubic spirals is largely analytical and is usually successful in finding a solution if it exists, whereas the algorithm for finding rational cubic spirals is entirely numerical and much more subject to discretization errors. (To recover those cases the grid size could be refined.) A hybrid algorithm may be considered that started with the cubic solutions and worked forward, but in our opinion it would not be worth the effort or computational expense. (Using this technique, it would obviously be possible to fix only cases where an analogous cubic exists. There are surely similar missed rational cubic spirals where cubic spirals do not exist.)

5.2.2. Necessary conditions for cubic spirals

The dark hyperbolas shown in Figs. 7–10 provide a lower boundary of the region in which any spirals may be located, as indicated by Inequality (8) and shown in Fig. 3. The horizontal and vertical lines split (K_0, K_1) space into four regions, of which, the upper right and lower left contain cubics (possibly

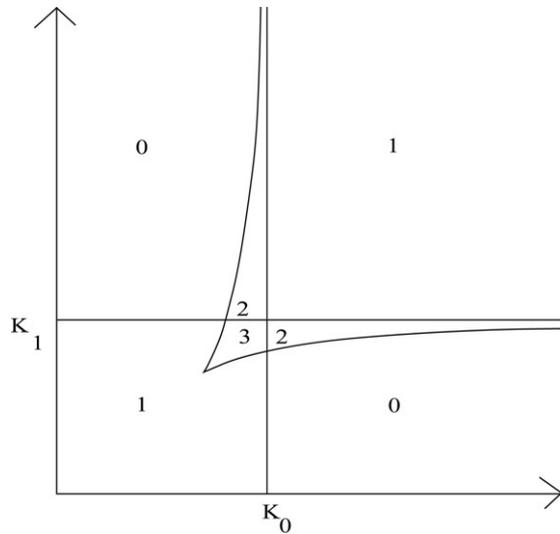


Fig. 11. Regions in (K_0, K_1) space with numbers showing multiplicity of cubics [3].

spirals) for a given (K_0, K_1) pair. The remaining curve (which has a cusp in it and limits towards the horizontal and vertical lines) cuts off a small region in which cubic spirals might also exist, because there are actually multiple cubics in that region for each (K_0, K_1) pair, as shown in Fig. 11. For example, for the conditions $\phi_0 = 0.3$, $\phi_1 = 0.7$, $K_0 = 0.3$, and $K_1 = 3.2$, there are three cubics. However, for $\phi_0 = 0.3$, $\phi_1 = 0.7$, $K_0 = 0.2$, and $K_1 = 4$, there are no cubics. (Comparison of the diagrams in Figs. 11 and 8 confirm this.) The remaining areas (upper left and lower right) will not result in cubics. For a further discussion see [3].

5.2.3. Insight into when the curvature spikes at the end

In the cases when both a cubic spiral and a rational cubic spiral exist and when the rational cubic spiral has a very steep curvature near $t = 1$ it is usually the case that the cubic spiral also has a very steep curvature near $t = 1$. In fact, upon studying this cubic, it usually also has a zero derivative at some value of t just slightly greater than 1. thus it has a spike in its curvature plot, but the top of the spike lies outside of $[0, 1]$.

In many of these cases, the cubic spirals are such that if the end curvatures are perturbed, the result is a curvature pair for which no cubic spiral exists and a spike occurs in the curvature plot near $t = 1$. This happens not only for small perturbations but also for many larger perturbations of the end curvatures as well. Thus, there is a large number of curvature pairs that occur in the search grid for which spikes occur near $t = 1$ in the cubics. To the extent the same phenomenon exists in rational cubics, the adaptation of the measure to append points near $t = 1$ is essential for the proper execution of the algorithm. To a lesser extent, a similar phenomenon occurs near $t = 0$.

5.2.4. Optimization of cubics versus rational cubics

It may be of interest to compare the values of the measure M for cubic spirals to those of rational cubic spirals in the cases that both exist for a given ϕ_0, ϕ_1, K_0, K_1 . Out of the 156 000 test cases ran, we found 20 309 cases where both cubic spirals

and rational cubic spirals existed. In 20 255 of these cases, the rational cubic spiral had a higher value of M , by a relative increase of about 5.67 on average. This is due to primarily to the fact the minimum slope of the curvature is larger. The other 54 cases were due to the fact that the search algorithm does not start with the cubic. But the resulting relative increase in the measure of the cubics over the rational cubics in this case was only about 3.38 on average.

We also experimented on this set of 20 309 spirals and computed an approximation to the maximum slope of the curvature. This is an alternate way of measuring the quality of a spiral, the smaller the slope, the better the spiral. Our algorithm did not try to optimize this. But it was found that for 13 507 cases the maximum approximate slope of the cubic was higher (worse) than the maximum approximate slope of the rational cubic by a relative increase on average of of about 0.42. For the other 6802 cases, the slope of the rational cubic was higher (worse) than that of the cubic by a relative increase on average of about 0.18. These are very slight differences in both cases indicating that the optimization of M does quite well in the majority of cases in this regard.

The measure M and the maximum slope of curvature are just two of many possible measures that might be used to judge the quality of a curve, but both are general purpose and appropriate for spirals. It should be noted that once a spiral has been found, it lies in the bounding crescent, and these bounding crescents are frequently even slimmer than the ones shown in Fig. 6. Thus, in the majority of cases, the slight differences in measures discussed in this section do not actually result in a perceptible visual improvement in the spiral.

5.3. Interesting spirals

The (K_0, K_1) pairs for which K_0 is near zero and K_1 is orders of magnitude larger are not as likely to be of interest in design settings, because they essentially mimic long lines with small hooks attached. Sometimes these hooks are too small to even be visible. While it seems easier to produce this kind of curve with rational spirals than with cubic spirals, this case is much more easily designed by using other shapes or piecewise curves. As such, it's not useful to investigate them for non-theoretical purposes.

Of more interest are the (K_0, K_1) pairs for which K_0 is somewhat larger than zero and K_1 is roughly of the same order of magnitude as K_0 . These are the spirals that come closest to equality in Inequality (8). For example, see Fig. 10, which shows both where cubic spirals exist and where rational cubic spirals exist in (K_0, K_1) space for $\phi_0 = 0.9$ and $\phi_1 = 1.4$. Note that the squares do not get as close to the bounding hyperbola curve as the diamonds. This illustrates a typical case where rational cubics offer greater flexibility and utility for design applications.

6. Conclusions

The algorithm presented is quite fast, easily running 100 cases in under a second on an average laptop computer. Thus,

it is quite feasible to design a fast, robust algorithm to find rational cubic spirals numerically. However, proper care must be taken to use an appropriate measure for optimization and to apply appropriate constraints so that the algorithm is not too susceptible to numerical errors arising from various typical curvature behaviors.

Acknowledgments

The authors would like to thank the referees for taking the time to read over the original draft and for making many valuable comments that greatly improved this paper.

References

- [1] Christoph Baumgarten, Gerald Farin. Approximation of logarithmic spirals. *Computer Aided Geometric Design* 1997;14:515–32.
- [2] Burchard HG, Ayers JA, Frey WH, Sapidis NS. Approximation with aesthetic constraints. In: Sapidis NS, editor. *Designing fair curves and surfaces*. Philadelphia: SIAM; 1994. p. 3–28.
- [3] Carl deBoor, Klaus Höllig, Malcolm Sabin. High accuracy geometric Hermite interpolation. *Computer Aided Geometric Design* 1987;4: 269–78.
- [4] Donna Dietz, Bruce Piper. Interpolation with cubic spirals. *Computer Aided Geometric Design* 2004;21:165–80.
- [5] Donna Dietz. *Convex Cubic Spirals*, Ph.D. dissertation. Troy (NY): Rensselaer Polytechnic Institute; 2002.
- [6] Gerald Farin. *Curves and surfaces for computer-aided geometric design*. New York: Academic Press; 1996.
- [7] Gerald Farin. In: Peters AK, editor. *NURBS, from projective geometry to practical use*. 2nd ed. MA: (Nattick); 1999 ISBN:1-56881-084-9.
- [8] Gibreel GM, Easa SM, Hassan Y, El-Dimeery IA. State of the art of highway geometric design consistency. *Journal of Transportation Engineering* 1999;125:305–13.
- [9] Goodman TNT, Meek DS. Planar interpolation with a pair of rational spirals. *Journal of Computational and Applied Mathematics* 2007;201: 112–27.
- [10] Guggenheimer HW. *Differential geometry*. New York: Dover; 1963.
- [11] Meek DS, Walton DJ. Planar spirals that match G^2 hermite data. *Computer Aided Geometric Design* 1998;15:103–26.
- [12] Monagan MB, Geddes KO, Heal KM, George Labahn, Vorkoetter SM, James McCarron, Paul DeMarco. *Maple 10 programming guide*. Maplesoft. Waterloo ON (Canada); 2005.
- [13] Walton DJ, Meek DS. A planar cubic Bézier spiral. *Journal of Computational and Applied Mathematics* 1996;72:85–100.