An analysis of cubic approximation schemes for conic sections

Michael Floater
SINTEF
P.O. Box 124, Blindern
0314 Oslo
Norway

May 1994, Revised April 1995 and January 1996

Abstract. In this paper a piece of a conic section is approximated by a cubic or piecewise cubic polynomial. The main tool is to define the two inner control points of the cubic as an affine combination, defined by $\lambda \in [0,1]$, of two control points of the conic. If λ is taken to depend on the weight w of the latter, a function $\lambda(w)$ results which is used to distinguish between different algorithms and to analyze their properties. One of the approximations is a piecewise cubic having G^4 continuity at the break points.

§1. Introduction

In this paper we approximate a segment of a conic section, represented as a rational quadratic Bézier curve, by cubic and piecewise cubic polynomials.

It is often desirable in geometric modelling to convert geometry from one representation to another and one kind of conversion which occurs frequently is from conic sections and surfaces like spheres, tori, cylinders and cones to polynomials [11].

By treating a conic like any other planar curve, a variety of standard algorithms for approximation [1] could be applied. The point in this paper is to exploit the particular structure of the conic in order to construct particularly good approximations.

The basic idea is to approximate the conic, described by control points $\mathbf{p}_0$, $\mathbf{p}_1$, $\mathbf{p}_2 \in \mathbb{R}^2$ and a weight $w > 0$, by a cubic Bézier curve with control points

$$\mathbf{p}_0,\ (1-\lambda)\mathbf{p}_0 + \lambda\mathbf{p}_1,\ (1-\lambda)\mathbf{p}_2 + \lambda\mathbf{p}_1,\ \mathbf{p}_2,$$

and $\lambda \in [0,1]$. By letting λ depend on w, we can characterize and analyze a number of approximations in terms of the function $\lambda(w)$.

Among the curve approximations which fit into this framework is the geometric Hermite interpolation for planar curves due to de Boor, Höllig, and Sabin [2]. Another is a mid-point Hermite interpolation similar to that presented by Dokken, Dæhlen, Lyche, and Mørken [5] for circular arcs.

When approximating the conic with several cubic pieces, i.e with a cubic spline, one can again characterize the approximating curve by a function $\lambda(w)$ provided each cubic piece is defined in the same way for each piece of the conic. The break points are best chosen by recursively subdividing the conic first into two subcurves, then four, etc. in such a way that each piece has the same weight. In practice only one or two subdivisions will be necessary before the approximation error is very small.

By choosing λ correctly, we will obtain both a classical C^2 cubic spline and a G^4 cubic spline. In this paper a curve will be said to have order of continuity k, written G^k, if it is C^k when reparametrized with respect to arc length [9]. Both splines are constructed explicitly without the need to solve a tridiagonal linear system.

The error of the G^4 cubic spline is shown to be $O(2^{-6r})$, where r is the level of subdivision (i.e. the spline has 2^r segments), making it apparently optimal in terms of both convergence order and smoothness. On the other hand the error is larger than when repeatedly using either of the Hermite interpolations mentioned earlier. Nevertheless the relative difference between the errors is small since all three are $O(2^{-6r})$. So one might consider using the G^4 cubic spline as an approximation to a conic which has both a very small error and a high degree of continuity.

In a recent paper by the author [8], conic sections were approximated by polynomial splines of any odd degree n. The approximation consists of geometric Hermite interpolation and the order of approximation is $O(h^{2n})$ where h is the maximum length of parameter interval. The spline segments join with G^{n-1} continuity. A related tensor-product spline surface was shown to approximate a class of tensor-product rational Bézier surfaces (essentially tensor-products of conic sections) in a similar way. In classical spline approximation the optimal approximation order and optimal smoothness are $O(h^{n+1})$ and C^{n-1} respectively [1]. Therefore the approximation in [8] is a considerable improvement on constructing a parametric spline approximation by approximating each component separately. Even the approximation using only a single polynomial segment has a very small error.

This and previous work on parametric approximation for general planar curves [2,3,14,15] suggest that order $2n$ is optimal when approximating with a polynomial of degree n. When using rational polynomials even higher orders of approximation have been found by Schaback [16] and Degen [4].

But what is the optimal order of geometric continuity? It is certainly not G^{n-1} since Schaback [15] constructed, using a shooting technique, an $O(h^4)$, G^2 quadratic spline interpolation through points satisfying a convexity condition.

By counting the number of degrees of freedom of a parametric polynomial, one is led to conjecture that the optimal order of continuity should be G^{2n-2} [13]. The existence of a G^4 cubic spline approximation for conic sections strengthens this conjecture.

This paper is organised in the following way. In Section 2, the approximation using a single cubic piece is defined and three Hermite interpolations are derived. In Section 3 the subdivision scheme is defined and it is proved that all the piecewise cubic approximations defined by $\lambda(w)$ and the subdivision scheme are C^1 and G^2. In Section 4, the two particular spline approximations, C^2 and G^4, are constructed. The order of convergence of the approximation error is studied in Section 5 together with a classification in terms of $\lambda(1)$ and $\lambda'(1)$. Numerical examples are provided in Section 6.

All curve approximations and error bounds discussed here can be extended to tensor product surfaces in the same way as [8].

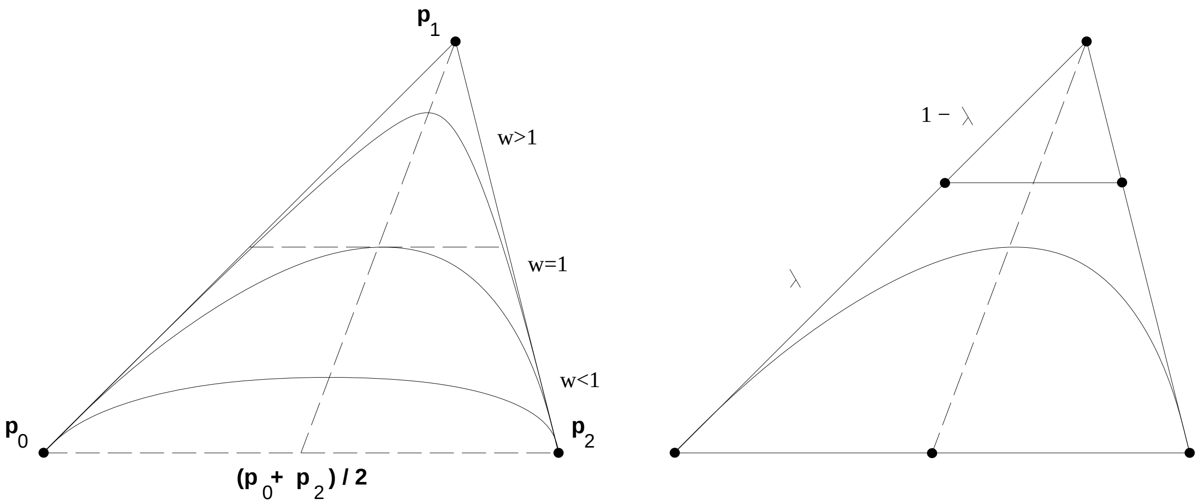

Figure 1. The conic $\mathbf{r}$ in the cases $0<w<1$, $w=1$ and $w>1$ and the approximation $\mathbf{q}$.

§2. Approximation by a single cubic

Consider the rational quadratic Bézier curve, in standard form,

$$\mathbf{r}(t) = \frac{B_{0,2}(t)\mathbf{p}_0 + B_{1,2}(t)w\mathbf{p}_1 + B_{2,2}(t)\mathbf{p}_2}{B_{0,2}(t) + B_{1,2}(t)w + B_{2,2}(t)} \tag{1}$$

where $\mathbf{p}_0$, $\mathbf{p}_1$, $\mathbf{p}_2 \in \mathbb{R}^2$ are the control points, $w > 0$ and $B_{i,n}(t) = \sum_{i=0}^{n} \binom{n}{i} t^i (1-t)^{n-i}$, are the Bernstein basis functions, and t is in the range [0,1]. It is well known [6,12] that $\mathbf{r}$ is an ellipse when $w < 1$, a parabola when $w = 1$ and a hyperbola when $w > 1$; see Figure 1. We are interested in the cubic approximation to $\mathbf{r}$ defined as

$$\mathbf{q}(t) = B_{0,3}(t)\mathbf{p}_0 + B_{1,3}(t)((1-\lambda)\mathbf{p}_0 + \lambda\mathbf{p}_1) + B_{2,3}(t)((1-\lambda)\mathbf{p}_2 + \lambda\mathbf{p}_1) + B_{3,3}(t)\mathbf{p}_2, \tag{2}$$

where $\lambda > 0$ is to be determined. Note that $\mathbf{q}$ and $\mathbf{r}$ meet tangentially at $t = 0, 1$. We are most interested in values of λ less than unity. Indeed, a well known theorem [10] states that if the control polygon of a Bézier curve is convex, then so is the curve. One can see from this that if $\lambda \le 1$, $\mathbf{q}$ is convex. Conversely, if $\lambda > 1$, the curvature of $\mathbf{q}$ at the end points has the opposite sign and $\mathbf{q}$ is no longer convex. In summary $\mathbf{q}$ is a convexity-preserving approximation to $\mathbf{r}$ if and only if $\lambda \le 1$. By the same reasoning, one can also see that $\mathbf{q}$ lies completely inside $\triangle\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$ if and only if $\lambda \le 1$.

We want to think of λ as a function of w; as one varies w, we wish to vary λ accordingly so that the Hausdorff distance $d_H(\mathbf{q}, \mathbf{r})$ between $\mathbf{q}$ and $\mathbf{r}$ is small. Recall that the Hausdorff distance [3] is defined as

$$d_H(\mathbf{q}, \mathbf{r}) = \max\left(\max_{t\in[0,1]} \min_{s\in[0,1]} |\mathbf{q}(s) - \mathbf{r}(t)|, \max_{s\in[0,1]} \min_{t\in[0,1]} |\mathbf{q}(s) - \mathbf{r}(t)|, \right).$$

In this simple framework, we will see how a number of cubic approximations can be represented uniquely as a function $\lambda(w)$. We can compare them by simply comparing the different λ functions.

An important tool for the analysis of the approximation error is the implicit form of $\mathbf{r}$. This idea was first used to study circle approximations by Dokken et al. [5] and has since been generalised in [8]. We can write any point $(x, y) \in \mathbb{R}^2$ in terms of barycentric coordinates τ_0, τ_1, τ_2, where $\tau_0 + \tau_1 + \tau_2 = 1$, with respect to the triangle $\triangle \mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$:

$$(x, y) = \tau_0\mathbf{p}_0 + \tau_1\mathbf{p}_1 + \tau_2\mathbf{p}_2.$$

The implicit form of $\mathbf{r}$ [12] can be defined as $f_{\mathbf{r}}(\mathbf{r}(t)) = 0$, where $f_{\mathbf{r}} : \mathbb{R}^2 \to \mathbb{R}$ is defined as

$$f_{\mathbf{r}}(x, y) = \tau_1^2 - 4w^2\tau_0\tau_2. \tag{3}$$

The function $f_{\mathbf{r}}$ is unique up to a scaling. We expect that if $f_{\mathbf{r}}(\mathbf{q})$ is small then $\mathbf{q}$ is close to $\mathbf{r}$. Indeed, the following theorem is derived from the form of $f_{\mathbf{r}}$ and its derivatives as a Bézier triangle and was proved in [8].

Theorem 1. *Suppose that $\mathbf{p} : [0, 1] \to \mathbb{R}^2$ is any regular curve which lies entirely inside the (closed) triangle $\triangle \mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$ and such that $\mathbf{p}(0) = \mathbf{p}_0$ and $\mathbf{p}(1) = \mathbf{p}_2$. Then*

$$d_H(\mathbf{p}, \mathbf{r}) \leq \frac{1}{4} \max\left(\frac{1}{w^2}, 1\right) \max_{t\in[0,1]} |f_{\mathbf{r}}(\mathbf{p}(t))| \, |\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2|.$$

Now we compute $f_{\mathbf{r}}(\mathbf{q}(t))$, a polynomial of degree six.

Proposition 2.

$$f_{\mathbf{r}}(\mathbf{q}(t)) = g_{w,\lambda}(t) = (1 - t)^2t^2(A(1 - t)^2 + B(1 - t)t + At^2),$$

where

$$A = 9\lambda^2 - 12w^2(1 - \lambda), \qquad \text{and} \qquad B = 18\lambda^2 - 4w^2(9(1 - \lambda)^2 + 1).$$

Proof. The barycentric coordinates of $\mathbf{q}$ are

$$\tau_0 = (1 - t + 3(1 - \lambda)t)(1 - t)^2, \qquad \tau_1 = 3\lambda(1 - t)t, \qquad \tau_2 = (3(1 - \lambda)(1 - t) + t)t^2.$$

These can then be substituted into (3). ◁

Notice that the polynomial $f_{\mathbf{r}}(\mathbf{q}(t))$ only depends on w and λ; it is independent of $\mathbf{p}_0$, $\mathbf{p}_1$, and $\mathbf{p}_2$.

We are now in a position to consider some geometric Hermite interpolations. The first thing to notice is that if $\lambda = \frac{2}{3}$ then $\mathbf{q}$ is merely the parabola

$$\mathbf{q}(t) = B_{0,2}(t)\mathbf{p}_0 + B_{1,2}(t)\mathbf{p}_1 + B_{2,2}(t)\mathbf{p}_2,$$

with degree raised to three. So $\mathbf{q}$ is quadratic and this is the idea behind the following simple approximation [7].

Proposition 3. *If λ is the constant function $\frac{2}{3}$, then $\mathbf{q}$ is a G^1 quadratic Hermite approximation to $\mathbf{r}$.*

Secondly, consider a more advanced example. We wish to find $\lambda(w)$ corresponding to the cubic Hermite approximation introduced for general curves by de Boor, Höllig & Sabin [2] in which tangent and curvature are matched at both ends.

Proposition 4. *Set* $\lambda = \frac{2}{3}w(\sqrt{w^2+3} - w)$. *Then* $\mathbf{q}$ *is a* G^2 *cubic Hermite approximation to* $\mathbf{r}$.

Proof. There are two ways to prove this. We could find λ in order that the curvatures of $\mathbf{q}$ and $\mathbf{r}$ agree at the endpoints by explicitly computing the curvatures there. But an easier proof is simply to demand that $f_{\mathbf{r}}(\mathbf{q}(t))$ has zeros at $t = 0, 1$ of multiplicity 2. Using this latter approach we see from Proposition 2 that the requirement is that

$$A = 9\lambda^2 - 12w^2(1-\lambda) = 0.$$

The two solutions to this quadratic equation are

$$\lambda = \frac{2}{3}w(\pm\sqrt{w^2+3} - w),$$

and the one for which λ lies in the range $(0, 1)$ is the one stated. ◁

Notice that there are two possible solutions for λ even though one of them is irrelevant. This ambiguity reflects the nonlinearity and nonuniqueness inherent in high order geometric approximations such as this one. In the case of the C^2 cubic spline (Section 4), λ is the unique solution to a linear equation.

The next Hermite approximation was found for circular arcs by Dokken et al. [5]. The approximating cubic has tangential contact at $t = 0, 1/2, 1$. Here it is generalised to conic sections. It is also a special case of the mid-point approximation of general odd degree n described in [8]. The form of λ in this case is particularly concise. Again there are two solutions but one of them is of no use.

Proposition 5. *Set* $\lambda = \frac{4}{3}w/(1+w)$. *Then* $\mathbf{q}$ *is a* G^1 *cubic Hermite mid-point approximation to* $\mathbf{r}$.

Proof. The requirement is that $f_{\mathbf{r}}(\mathbf{q}(t))$ should have a zero of multiplicity two at $t = 1/2$. Since this polynomial is symmetric it suffices to set $f_{\mathbf{r}}(\mathbf{q}(\frac{1}{2})) = 0$. This implies that $2A + B = 0$ and therefore

$$36\lambda^2 - 24w^2(1-\lambda) - 4w^2(9(1-\lambda)^2 + 1) = 0,$$

or

$$9\lambda^2 - w^2(4 - 3\lambda)^2 = 0.$$

So

$$3\lambda = \pm w(4 - 3\lambda),$$

which implies that either

$$\lambda = \frac{4}{3}w/(1+w) \qquad \text{or} \qquad \lambda = -\frac{4}{3}w/(w-1),$$

and the second solution is hardly acceptable when w is close or equal to 1. ◁

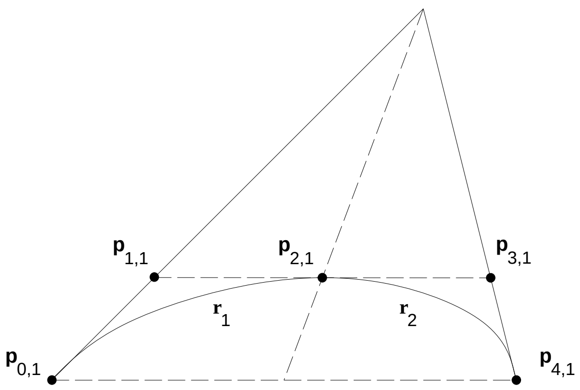

Fig. 2. The first subdivision of $\mathbf{r}$ (with $w<1$).

§3. Approximation by piecewise cubics

We now wish to consider approximating the conic with a piecewise cubic curve. The order of continuity between the pieces will be raised if the break points are well chosen. We will choose break points in a natural way by recursively subdividing $\mathbf{r}$. At each step each piece of the conic is divided at its mid-point and each subcurve is re-expressed in the standard form (1). The error of approximation will get small very quickly after only one or two subdivisions.

First of all we set $\mathbf{q}_0 = \mathbf{q}$, $w_0 = w$, and $\mathbf{p}_{i,0} = \mathbf{p}_i$. Consider the first subdivision of $\mathbf{r}$. Letting $\mathbf{r}_1$ be the subcurve $\mathbf{r}|_{0\le t\le 1/2}$ and $\mathbf{r}_2$ be the subcurve $\mathbf{r}|_{1/2\le t\le 1}$ one finds

$$\mathbf{r}_1(t) = \frac{B_0(t)\mathbf{p}_{0,1} + B_1(t)w_1\mathbf{p}_{1,1} + B_2(t)\mathbf{p}_{2,1}}{B_0(t) + B_1(t)w_1 + B_2(t)}$$

and

$$\mathbf{r}_2(t) = \frac{B_0(t)\mathbf{p}_{2,1} + B_1(t)w_1\mathbf{p}_{3,1} + B_2(t)\mathbf{p}_{4,1}}{B_0(t) + B_1(t)w_1 + B_2(t)}$$

for $t \in [0,1]$ where $w_1 = \sqrt{(1+w_0)/2}$ and

$$\begin{aligned}
\mathbf{p}_{0,1} =& \mathbf{p}_0,\\
\mathbf{p}_{1,1} =& (\mathbf{p}_0 + w_0\mathbf{p}_1)/(1+w_0),\\
\mathbf{p}_{2,1} =& (\mathbf{p}_0 + 2w_0\mathbf{p}_1 + \mathbf{p}_2)/2(1+w_0),\\
\mathbf{p}_{3,1} =& (w_0\mathbf{p}_1 + \mathbf{p}_2)/(1+w_0),\\
\mathbf{p}_{4,1} =& \mathbf{p}_2;
\end{aligned}$$

see Figure 2. Note that the parametrizations of the two subcurves are related nonlinearly with that of $\mathbf{r}$. But this is of no concern as we are only considering approximation schemes which are independent of the parametrization of $\mathbf{r}$.

Corresponding to any chosen function λ, one can now approximate $\mathbf{r}_1$ and $\mathbf{r}_2$ by

$$\begin{aligned}
\mathbf{q}_{1,1}(t) =& B_{0,3}(t)\mathbf{p}_{0,1} + B_{1,3}(t)((1-\lambda(w_1))\mathbf{p}_{0,1} + \lambda(w_1))\mathbf{p}_{1,1}\\
&+ B_{2,3}(t)((1-\lambda(w_1))\mathbf{p}_{2,1} + \lambda(w_1))\mathbf{p}_{1,1} + B_{3,3}(t)\mathbf{p}_{2,1},
\end{aligned}$$

and

$$\begin{aligned}\mathbf{q}_{2,1}(t) =& B_{0,3}(t)\mathbf{p}_{2,1} + B_{1,3}(t)((1-\lambda(w_1))\mathbf{p}_{2,1} + \lambda(w_1))\mathbf{p}_{3,1} \\ &+ B_{2,3}(t)((1-\lambda(w_1))\mathbf{p}_{4,1} + \lambda(w_1))\mathbf{p}_{3,1} + B_{3,3}(t)\mathbf{p}_{4,1},\end{aligned}$$

respectively. The Hausdorff distance between $\mathbf{r}$ and the approximation $\mathbf{q}_1$ consisting of the curve segments $\mathbf{q}_{1,1}$ and $\mathbf{q}_{2,1}$, is then clearly bounded by the greater of $d_H(\mathbf{q}_1, \mathbf{r}_1)$ and $d_H(\mathbf{q}_2, \mathbf{r}_2)$. Continuing in this way one finds the following scheme.

Subdivision scheme.

For $r = 1, 2, \ldots$ and $i = 0, \ldots, 2^{r-1} - 1$, define

$$\begin{aligned}\mathbf{p}_{4i,r} =& \mathbf{p}_{2i,r-1}, \\ \mathbf{p}_{4i+1,r} =& (\mathbf{p}_{2i,r-1} + w_{r-1}\mathbf{p}_{2i+1,r-1})/(1+w_{r-1}), \\ \mathbf{p}_{4i+2,r} =& (\mathbf{p}_{2i,r-1} + 2w_{r-1}\mathbf{p}_{2i+1,r-1} + \mathbf{p}_{2i+2,r-1})/2(1+w_{r-1}), \\ \mathbf{p}_{4i+3,r} =& (w_{r-1}\mathbf{p}_{2i+1,r-1} + \mathbf{p}_{2i+2,r-1})/(1+w_{r-1}), \\ \mathbf{p}_{4i+4,r} =& \mathbf{p}_{2i+2,r-1},\end{aligned} \tag{4}$$

and let $w_r = \sqrt{(1 + w_{r-1})/2}$.

For each r, the curve $\mathbf{r}$ consists of the 2^r segments

$$\frac{B_{0,2}(t)\mathbf{p}_{2i,r} + B_{1,2}(t)w_r\mathbf{p}_{2i+1,r} + B_{2,2}(t)\mathbf{p}_{2i+2,r}}{B_{0,2}(t) + B_{1,2}(t)w_r + B_{2,2}(t)} \tag{5}$$

for $i = 0, \ldots, 2^r - 1$. We define the r-th spline approximation $\mathbf{q}_r$, to be the piecewise polynomial curve consisting of the 2^r associated segments

$$\begin{aligned}&B_{0,3}(t)\mathbf{p}_{2i,r} + B_{1,3}(t)((1-\lambda(w_r))\mathbf{p}_{2i,r} + \lambda(w_r)\mathbf{p}_{2i+1,r}) \\ &\quad + B_{2,3}(t)((1-\lambda(w_r))\mathbf{p}_{2i+2,r} + \lambda(w_r)\mathbf{p}_{2i+1,r}) + B_{3,3}(t)\mathbf{p}_{2i+2,r}.\end{aligned} \tag{6}$$

There is a natural symmetry in the subdivision scheme which can be understood by considering the case when $\mathbf{r}$ is a circular arc with $|\mathbf{p}_0 - \mathbf{p}_1| = |\mathbf{p}_2 - \mathbf{p}_1|$ and $w = \cos(\theta/2)$, where θ is the angle subtended by the arc (equal to $\pi - L\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$). In this case the subdivision of $\mathbf{r}$ is uniform with respect to arc length and the angle it subtends.

As one continues to subdivide $\mathbf{r}$, it is evident that $w_r \to 1$ as $r \to \infty$ where r is the level of subdivision. For this reason, the order of approximation of any scheme defined by the function λ depends only on λ and its derivatives at the value 1. This is discussed later.

We will now observe that whatever the function $\lambda(w)$, every spline approximation $\mathbf{q}_r$ is both C^1 and G^2. This requires a "connection" lemma, proved in [7], which relates the five control points of every consecutive pair of segments of $\mathbf{r}$.

Lemma 6. *For all subdivision levels $r \geq 1$ and all $i = 0, \ldots, 2^r - 2$, one has*

$$\mathbf{p}_{2i+1,r} - 2\mathbf{p}_{2i+2,r} + \mathbf{p}_{2i+3,r} = 0.$$

and

$$\mathbf{p}_{2i+4,r} - \mathbf{p}_{2i,r} = 2w_r^2(\mathbf{p}_{2i+3,r} - \mathbf{p}_{2i+1,r}).$$

For example, when $r = 2$, $\mathbf{r}$ is represented by four rational pieces with the same weight w_2 and control points $\{\mathbf{p}_{0,2}, \mathbf{p}_{1,2}, \mathbf{p}_{2,2}\}$, $\{\mathbf{p}_{2,2}, \mathbf{p}_{3,2}, \mathbf{p}_{4,2}\}$, $\{\mathbf{p}_{4,2}, \mathbf{p}_{5,2}, \mathbf{p}_{6,2}\}$ and $\{\mathbf{p}_{6,2}, \mathbf{p}_{7,2}, \mathbf{p}_{8,2}\}$. The first identity in Lemma 6 relates the control points $\{\mathbf{p}_{1,2}, \mathbf{p}_{2,2}, \mathbf{p}_{3,2}\}$, $\{\mathbf{p}_{3,2}, \mathbf{p}_{4,2}, \mathbf{p}_{5,2}\}$ and $\{\mathbf{p}_{5,2}, \mathbf{p}_{6,2}, \mathbf{p}_{7,2}\}$ in an identical way and the second $\{\mathbf{p}_{0,2}, \mathbf{p}_{1,2}, \mathbf{p}_{3,2}, \mathbf{p}_{4,2}\}$, $\{\mathbf{p}_{2,2}, \mathbf{p}_{3,2}, \mathbf{p}_{5,2}, \mathbf{p}_{6,2}\}$ and $\{\mathbf{p}_{4,2}, \mathbf{p}_{5,2}, \mathbf{p}_{7,2}, \mathbf{p}_{8,2}\}$ also in an identical way.

Proposition 7. *For any $r \geq 1$, the curve $\mathbf{q}_r$ is C^1 and G^2 and meets $\mathbf{r}$ tangentially at the ends of every polynomial segment.*

Proof. Since the proof of continuity depends only on the two connection identities in Lemma 6 and these hold for all r and i, we can reduce the number of subscripts by defining

$$\mathbf{p}(t) = B_{0,3}(t)\mathbf{q}_0 + B_{1,3}(t)((1-\lambda)\mathbf{q}_0 + \lambda\mathbf{q}_1) + B_{2,3}(t)((1-\lambda)\mathbf{q}_2 + \lambda\mathbf{q}_1) + B_{3,3}(t)\mathbf{q}_2, \quad (7)$$

and

$$\mathbf{q}(t) = B_{0,3}(t)\mathbf{q}_2 + B_{1,3}(t)((1-\lambda)\mathbf{q}_2 + \lambda\mathbf{q}_3) + B_{2,3}(t)((1-\lambda)\mathbf{q}_4 + \lambda\mathbf{q}_3) + B_{3,3}(t)\mathbf{q}_4, \quad (8)$$

for given points $\mathbf{q}_i$. It is required to show that if

$$\mathbf{q}_1 - 2\mathbf{q}_2 + \mathbf{q}_3 = 0. \quad (9)$$

and

$$\mathbf{q}_4 - \mathbf{q}_0 = 2w^2(\mathbf{q}_3 - \mathbf{q}_1), \quad (10)$$

then $\mathbf{p}$ and $\mathbf{q}$ join with C^1 and G^2 continuity at $t = 1$ and $t = 0$ respectively.

Differentiating $\mathbf{p}$ and $\mathbf{q}$ yields

$$\mathbf{p}'(1) = 3\lambda(\mathbf{q}_2 - \mathbf{q}_1), \qquad \mathbf{q}'(0) = 3\lambda(\mathbf{q}_3 - \mathbf{q}_2), \quad (11)$$

and from (9) it follows that $\mathbf{p}'(1) = \mathbf{q}'(0)$, and this shows that the curves join with C^1 continuity. To prove G^2 continuity [9] it is necessary to show the existence of a scalar β_2 for which

$$\mathbf{q}''(0) = \beta_2\mathbf{p}'(1) + \mathbf{p}''(1).$$

Differentiating again yields

$$\mathbf{p}''(1) = 6((1-\lambda)\mathbf{q}_0 - \lambda\mathbf{q}_1 + (2\lambda - 1)\mathbf{q}_2), \quad (12)$$

$$\mathbf{q}''(0) = 6((2\lambda - 1)\mathbf{q}_2 - \lambda\mathbf{q}_3 + (1-\lambda)\mathbf{q}_4). \quad (13)$$

Therefore

$$\mathbf{q}''(0) - \mathbf{p}''(1) = 6((1-\lambda)(\mathbf{q}_4 - \mathbf{q}_0) - \lambda(\mathbf{q}_3 - \mathbf{q}_1)),$$

and so from (9) and (11) we find that β_2 does indeed exist and

$$\beta_2 = \frac{4}{\lambda}(2w^2(1-\lambda) - \lambda). \quad (14)$$

The fact that $\mathbf{q}_r$ touches $\mathbf{r}$ tangentially is obvious from the fact that each local approximation has tangential contact. ◁

In particular if $\mathbf{q}_r$ is a Hermite approximation, using any of the λ's defined in Propositions 3, 4, and 5, its continuity at the break points will be C^1 and G^2. This higher order of continuity is due to $\mathbf{r}$ being a conic and to the choice of break points.

§4. Cubic spline approximation

By constraining the function $\lambda(w)$ in such a way that each pair of curve segments of the r-th spline approximation $\mathbf{q}_r$ join with C^2 continuity, we will obtain a classical C^2 piecewise cubic interpolation without solving a tridiagonal linear system.

Proposition 8. *Let* $\lambda = 2w^2/(1+2w^2)$. *Then* $\mathbf{q}_r$ *is a* C^2 *cubic spline interpolation to* $\mathbf{r}$.

Proof. We just need to choose λ in such a way that $\beta_2 = 0$ in Proposition 7, in other words we require

$$2w^2(1-\lambda) - \lambda = 0.$$

This is a linear equation in λ, unlike in the geometric Hermite cases, and we obtain the unique solution

$$\lambda = \frac{2w^2}{1+2w^2},$$

as claimed. ◁

It has recently been suggested [13] that it should be possible in principle to approximate a planar curve by a piecewise cubic spline which has continuity G^4 between the segments. One would expect that there might have to be some constraints on the curve, for example, it might have to be convex. Conic sections are always convex.

We will now show that it is possible to constrain $\lambda(w)$ in such a way that each pair of curve segments of $\mathbf{q}_r$ join with G^4 continuity and thus obtain a G^4 cubic spline approximation for the conic.

It will be shown later that the order of approximation is $O(2^{-6r})$ which is higher than for the C^2 cubic spline. This provides striking evidence that arc-length continuity has a positive effect on the error and order of approximation. The equations are nonlinear in nature but again the algebra involved in solving them is simplified by the fact that any pair of adjacent segments can be considered in isolation. In fact the key to finding the G^4 cubic is simply to find the G^3 cubic.

Proposition 9. *Define* $\lambda = 2(6w^2+1-\sqrt{3w^2+1})/(12w^2+3)$. *Then* $\mathbf{q}_r$ *is a* G^3 *cubic spline approximation to* $\mathbf{r}$.

Proof. In analogy with Proposition 7 and Proposition 8, the task is to find λ such that the two subcurves $\mathbf{p}$ and $\mathbf{q}$ defined in (7) and (8) join with G^3 continuity provided that the connection equations (9) and (10) hold. This means that we must find constants β_1, β_2, β_3, such that $\mathbf{p}$ and $\mathbf{q}$ satisfy the identities

$$\begin{aligned} \mathbf{q}'(0) &= \beta_1 \mathbf{p}'(1), \\ \mathbf{q}''(0) &= \beta_2 \mathbf{p}'(1) + \beta_1^2 \mathbf{p}''(1), \\ \mathbf{q}'''(0) &= \beta_3 \mathbf{p}'(1) + 3\beta_1\beta_2 \mathbf{p}''(1) + \beta_1^3 \mathbf{p}'''(1). \end{aligned}$$

The coefficients of the derivatives of $\mathbf{p}$ are the elements of the second to fourth rows of the well known connection matrix which connects $\mathbf{p}$ and $\mathbf{q}$ with the desired degree of arc-length continuity [9].

The first and second derivatives were computed in (11), (12), and (13). The third derivatives are

$$\mathbf{p}'''(1) = 6(2-3\lambda)(\mathbf{q}_0 - \mathbf{q}_2), \qquad \text{and} \qquad \mathbf{q}'''(0) = 6(2-3\lambda)(\mathbf{q}_2 - \mathbf{q}_4). \tag{15}$$

It was observed earlier that β_1 and β_2 exist and that $\beta_1 = 1$ and

$$\beta_2 = \frac{4}{\lambda}(2w^2(1-\lambda) - \lambda).$$

So it remains to choose λ in order that β_3 should exist. Employing (9) and (10), one finds

$$\begin{aligned}\beta_3 \mathbf{p}'(1) =& \mathbf{q}'''(0) - \mathbf{p}'''(1) - 3\beta_2 \mathbf{p}''(1) \\ =& 6(2-3\lambda)(2(\mathbf{q}_2 - \mathbf{q}_0) - 4w^2(\mathbf{q}_2 - \mathbf{q}_1)) \\ &+ \frac{72}{\lambda}(2w^2(1-\lambda) - \lambda)(-(1-\lambda)(\mathbf{q}_2 - \mathbf{q}_0) + \lambda(\mathbf{q}_2 - \mathbf{q}_1)),\end{aligned}$$

and this reduces to

$$\beta_3 \mathbf{p}'(1) = C(\mathbf{q}_2 - \mathbf{q}_1) + D(\mathbf{q}_2 - \mathbf{q}_0), \tag{16}$$

where

$$C = 12\left(-2(2-3\lambda)w^2 - 6(2w^2(1-\lambda) - \lambda)\right),$$

$$D = 12\left((2-3\lambda)w^2 + \frac{6}{\lambda}(2w^2(1-\lambda) - \lambda)(1-\lambda)\right).$$

Since the vectors $\mathbf{q}_2 - \mathbf{q}_1$ and $\mathbf{q}_2 - \mathbf{q}_0$ are always linearly independent, β_3 exists if and only if $D = 0$. After a little more algebra we then find that

$$w^2 = \frac{(4-3\lambda)\lambda}{12(1-\lambda)^2}. \tag{17}$$

If instead we solve for λ in terms of w^2 we find two solutions

$$\lambda = 2(6w^2 + 1 \pm \sqrt{3w^2+1})/(12w^2+3).$$

If one takes the positive square root, λ is greater than 1 for $w < 4$ and this is highly undesirable. Thus we have to take the negative root and one can easily check that in that case, $0 < \lambda < 1$ for all $w > 0$. $\triangleleft$

So there is essentially only one G^3 cubic spline approximation. It turns out to have the property of also being G^4. This can be shown by affinely mapping each pair of conic pieces into a new pair which are symmetries of each other and observing that the corresponding cubic pieces share this symmetry. One can then use a symmetry argument as explained in [8]. However it is enlightening to use the machinery already developed and the connection matrix once again.

Proposition 10. *The G^3 cubic spline approximation defined by setting $\lambda = 2(6w^2 + 1 - \sqrt{3w^2+1})/(12w^2+3)$ is also G^4.*

Proof. Consider the next equation from the connection matrix:

$$\mathbf{q}^{(4)}(0) = \beta_4 \mathbf{p}'(1) + (3\beta_2^2 + 4\beta_1\beta_3)\mathbf{p}''(1) + 6\beta_1^2\beta_2\mathbf{p}'''(1) + \beta_1^4\mathbf{p}^{(4)}(1).$$

The first thing to do is to compute β_2. Substituting (17) into (14) one obtains

$$\beta_2 = \frac{2(3\lambda - 2)}{3(1-\lambda)}.$$

We also need β_3. Since $D = 0$ in (16) and $\mathbf{p}'(1) = 3\lambda(\mathbf{q}_2 - \mathbf{q}_1)$, it follows that $\beta_3 = C/3\lambda$. Using equation (17), this implies that $\beta_3 = 2(2-3\lambda)^2/3(1-\lambda)^2$ and so

$$\beta_3 = \tfrac{3}{2}\beta_2^2 > 0.$$

Then, since $\mathbf{q}^{(4)}(0) = \mathbf{p}^{(4)}(1) = 0$, one finds that

$$\begin{aligned}\beta_4\mathbf{p}'(1) &= -6\beta_2\mathbf{p}'''(1) - (3\beta_2^2 + 4\beta_3)\mathbf{p}''(1)\\ &= -3\beta_2(2\mathbf{p}'''(1) + 3\beta_2\mathbf{p}''(1)) = -\frac{36\lambda(3\lambda-2)\beta_2}{(1-\lambda)}(\mathbf{q}_2 - \mathbf{q}_1).\end{aligned}$$

and this shows that β_4 exists and therefore that the curve segments join with G^4 continuity. Notice also that

$$\beta_4 = -18\beta_2^2 < 0.$$

Since the algebra here only depends on (17), it is evident that the other, non-useful G^3 solution is also G^4. ◁

§5. Order of approximation

From now on it will be understood that $\lambda : (0,\infty) \to \mathbb{R}$ is a C^2 function. What is the approximation order of the various piecewise cubic approximations studied so far? Since $w_r \to 1$ as $r \to \infty$, one would expect that the approximation order only depends on the behaviour of λ in a neighbourhood of 1. The following theorem gives sufficient conditions for a range of approximation orders. For continuous $g : [0,1] \to \mathbb{R}$, let $\|g\|_\infty = \max_{t\in[0,1]} |g(t)|$.

Proposition 11. *If $0 < \lambda(1) < 1$, the spline approximation $\mathbf{q}_r$ is second order accurate as $r \to \infty$. If $\lambda(1) = \frac{2}{3}$ then the approximation is fourth order accurate. If in addition $\lambda'(1) = \frac{1}{3}$ then the approximation is sixth order accurate.*

Proof. If $0 < \lambda(1) < 1$, then, since $w_r \to 1$, we must have $0 < \lambda(w_r) < 1$ for large enough r. In that case Theorem 1 applies and due to the subdivision scheme, we have

$$d_H(\mathbf{q}_r, \mathbf{r}) \le \frac{1}{4}\max\left(\frac{1}{w_r^2}, 1\right)\|g_{w_r,\lambda(w_r)}\|_\infty \max_{i=0,\ldots,2^r-1} |\mathbf{p}_{2i,r} - 2\mathbf{p}_{2i+1,r} + \mathbf{p}_{2i+2,r}|. \tag{18}$$

It was shown in [7] that $\max_{i=0,\ldots,2^r-1} |\mathbf{p}_{2i,r} - 2\mathbf{p}_{2i+1,r} + \mathbf{p}_{2i+2,r}|$ is $O(2^{-2r})$ and it follows that $d_H(\mathbf{q}_r, \mathbf{r})$ is also $O(2^{-2r})$.

Now suppose that $\lambda(1) = \frac{2}{3}$. To show that $d_H(\mathbf{q}_r, \mathbf{r})$ is $O(2^{-4r})$ it is sufficient to demonstrate that $\|g_{w_r,\lambda(w_r)}\|_\infty$ is $O(2^{-2r})$. Since [7] the quantity $a_r = w_r - 1$ is $O(2^{-2r})$, this is the case if both A and B in Proposition 2 are $O(a)$ as functions of $a = w - 1$. The Taylor expansion for λ about 1 is

$$\lambda(w) = \lambda(1+a) = \frac{2}{3} + \lambda'(1)a + O(a^2),$$

and so expanding A and B in terms of a, we find

$$A = 8(3\lambda'(1) - 1)a + O(a^2), \qquad \text{and} \qquad B = 16(3\lambda'(1) - 1)a + O(a^2).$$

This shows that A and B are indeed $O(a)$ as $a \to 0$. Therefore the approximation in this case is fourth order accurate.

Lastly, if in addition $\lambda'(1) = \frac{1}{3}$ it again follows from the Taylor expansion that A and B are both $O(a^2)$ in which case $\|g_{w_r,\lambda(w_r)}\|_\infty$ is $O(2^{-4r})$. Hence the approximation is then sixth order accurate. ◁

We can apply Proposition 11 to obtain the approximation orders with respect to r of the particular cubic splines described in this paper. All of these have been proved before except for the G^4 cubic spline.

Proposition 12. *The geometric Hermite, mid-point Hermite and G^4 spline approximations are all sixth order accurate. The G^2 quadratic Hermite and C^2 cubic spline approximations are both fourth order accurate.*

Proof. Straightforward calculations show that $\lambda(1) = \frac{2}{3}$ in all five cases and $\lambda'(1) = \frac{1}{3}$ in the first three. ◁

In order to implement the approximation scheme, we require a stopping criterion. One such criterion is to stop when the upper bound (18) on the Hausdorff error is within a given tolerance. This requires computing $\|g_{w_r,\lambda(w_r)}\|_\infty$ or an upper bound on it. To this end consider the following.

Proposition 13. *If either $3B < -2A < 0$ or $3B > -2A > 0$, then*

$$\|g_{w,\lambda}\|_\infty = \max\left(\ |2A+B|/64, \quad 4|A|^3/27(2A-B)^2 \ \right).$$

Otherwise

$$\|g_{w,\lambda}\|_\infty = |2A+B|/64.$$

Proof. Letting $u = t - 1/2$ and $h(u) = g(t)$, we find

$$h = (1/4 - u^2)^2((2A+B)/4 + (2A-B)u^2).$$

Then

$$h'(u) = (1/4 - u^2)u(-(2A+B) + (1/2 - 6u^2)(2A-B)),$$

so the solutions to $h'(u) = 0$ are $u = 0, 1/2, u_1$ where $u_1^2 = (-2A-3B)/12(2A-B)$. In order that u_1 is real and that $|u_1| < 1/2$ (i.e. so that $t \in (0,1)$), it is required, after some manipulation, that either $3B < -2A < 0$ or $3B > -2A > 0$. The maximum of h occurs at either $u = 0$ or $u = u_1$ if u_1 exists. Substituting these two values into h we find

$$h(0) = (2A+B)/64$$

and

$$h(u_1) = 4A^3/27(2A-B)^2.$$

Taking the modulus of these values yields the result. ◁

Using Proposition 13, it is straightforward to calculate $\|g_{w,\lambda}\|_\infty$ for any given lambda function.

Is it possible to choose $\lambda''(1)$ such that the approximation is eighth order accurate? A precise answer to this seems to be quite delicate even though we expect it to be ‘no’. One would require a lower bound on the Hausdorff error. What we can say is the following, assuming λ is C^3.

Proposition 14. *If* $\lambda(1) = \frac{2}{3}$ *and* $\lambda'(1) = \frac{1}{3}$ *then* $2A - B - 4a^2 = O(|a|^3)$. *Therefore it is not possible to choose* $\lambda(1)$, $\lambda'(1)$, *and* $\lambda''(1)$ *so that* $\|g_{w_r,\lambda(w_r)}\|_\infty$ *is* $O(2^{-6r})$.

Proof. Since

$$\lambda(w) = \lambda(1+a) = \frac{2}{3} + \frac{1}{3}a + \frac{1}{2!}\lambda''(1)a^2 + O(|a|^3),$$

it follows that

$$A = (12\lambda''(1) + 5)a^2 + O(|a|^3) \qquad \text{and} \qquad B = (24\lambda''(1) + 6)a^2 + O(|a|^3),$$

and therefore $2A - B = 4a^2 + O(|a|^3)$. Thus no choice of $\lambda''(1)$ raises simultaneously the order of convergence of A and B. Yet it is clear from the proof of Proposition 11 that it is necessary that $\lambda(1) = \frac{2}{3}$ and $\lambda'(1) = \frac{1}{3}$ in order that A and B are $O(a^2)$. Hence $\|g_{w_r,\lambda(w_r)}\|_\infty$ can never be $O(2^{-6r})$. ◁

§6. Illustrations

Figure 3 shows the function $\lambda(w)$ and one can clearly see how for all the sixth order approximations, $\lambda(1) = 2/3$ and $\lambda'(1) = 1/3$. Meanwhile $\lambda'(1) \neq 1/3$ for the two fourth order ones.

Figures 4 and 5 show the polynomial $f_{\mathbf{r}}(\mathbf{q}(t))$ in two cases where $w < 1$ and $w > 1$. Note that the three sixth order approximations lie on the same side of $\mathbf{r}$ regardless of the value of w. Meanwhile the fourth order ones switch sides as w crosses the value 1. This is to be expected. The polynomial $g_{w,\lambda}(t)$ is $O(a^2)$ in the former case and only $O(a)$ in the latter. Therefore only in the latter case do the leading terms change sign when $a = w - 1$ changes sign.

Figures 6 and 7 show an example of $\mathbf{r}$ and the various local approximations $\mathbf{q}$ in two cases, $w < 1$ and $w > 1$. In Figure 6, the mid-point Hermite approximation cannot be seen since it appears to lie exactly on $\mathbf{r}$ at this scale. In Figure 7 the mid-point Hermite approximation is again a good one but one can see that it lies outside $\triangle\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$ near $\mathbf{p}_0$ and $\mathbf{p}_2$ and it is not convex. These are both consequences of the fact that $\lambda > 1$ for $w > 3$ in this case.

Figure 8 shows the curvature of the three sixth order approximations $\mathbf{q}_1$ (one subdivision) when $\mathbf{r}$ is a quadrant of a unit circle. The fact that the G^4 cubic spline is G^3 i.e. has smooth curvature, shows up clearly in sharp contrast to the other two. The symmetry of the approximations and the corresponding G^4 continuity of the G^4 spline are also clear in this case.

§7. References

1. de Boor C., *A Practical Guide to Splines*, Springer-Verlag, New York, 1978.
2. de Boor C., Höllig K., Sabin M., High accuracy geometric Hermite interpolation, Computer-Aided Geom. Design **4** (1987), 269–278.
3. Degen W., Best approximations of parametric curves by splines, in *Mathematical Methods in CAGD, T. Lyche & L. L. Schumaker (eds.), Academic Press, Boston* **2** (1992), 171–184.
4. Degen W., High accurate rational approximation of parametric curves, Computer-Aided Geom. Design **10** (1993), 293–313.
5. Dokken T., Dæhlen M., Lyche T., Mørken K., Good approximation of circles by curvature-continuous Bézier curves, Computer-Aided Geom. Design , **7** (1990), 33–41.
6. Farin G., *Curves and surfaces for computer aided geometric design*, Academic Press, San Diego, 1988.
7. Floater M. S., High order approximation of conic sections by quadratic splines, Computer-Aided Geom. Design **12** (1995), 617–637.
8. Floater M. S., An $O(h^{2n})$ Hermite approximation for conic sections, preprint.
9. Goodman T. N. T., Properties of β-splines, J. Approx. Th. **44** (1985), 132–153.
10. Goodman T. N. T., Shape preserving representations, in *Mathematical Methods in CAGD, T. Lyche & L. L. Schumaker (eds.), Academic Press, Boston* (1989), 333–351.
11. Hoschek J., Schneider F., Spline conversion for trimmed rational Bézier- and B-spline surfaces, Comput. Aided Design **22** (1990), 580–590.
12. Lee E., The rational Bézier representation for conics, in *Geometric Modeling: Algorithms and New Trends, G. Farin (ed.), SIAM, Philadelphia* (1987), 3–19.
13. Mørken K., private communication.
14. Mørken K., Scherer K., A general framework for high accuracy parametric interpolation, preprint.
15. Schaback R., Interpolation with piecewise quadratic visually C^2 Bézier polynomials, Computer-Aided Geom. Design **6** (1989), 219–233.
16. Schaback R., Planar curve interpolation by piecewise conics of arbitrary type, Computer-Aided Geom. Design **9** (1993), 373–389.

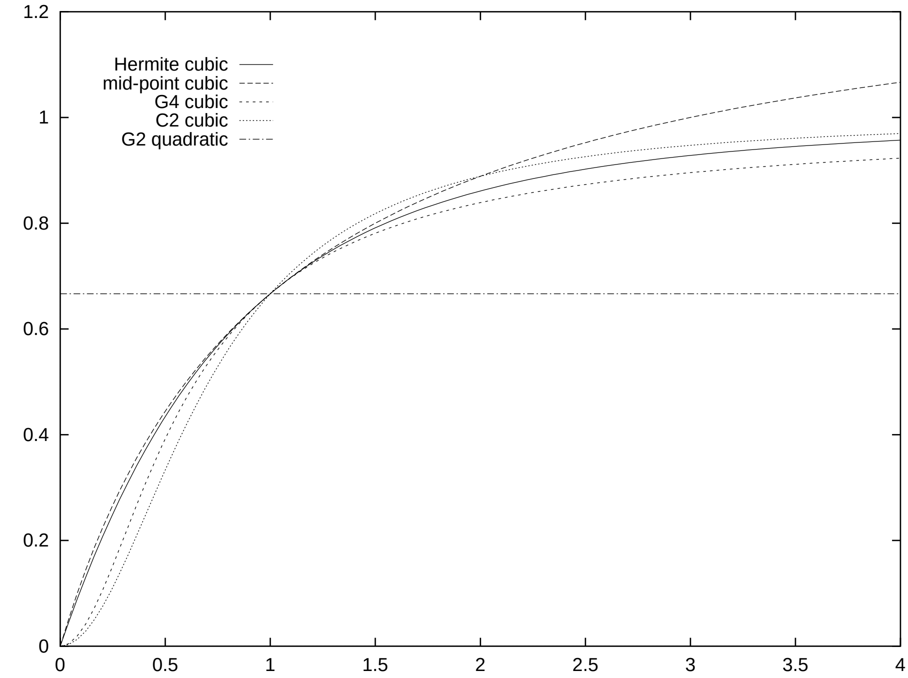

Figure 3. The function $\lambda(w)$ for $0<w<4$.

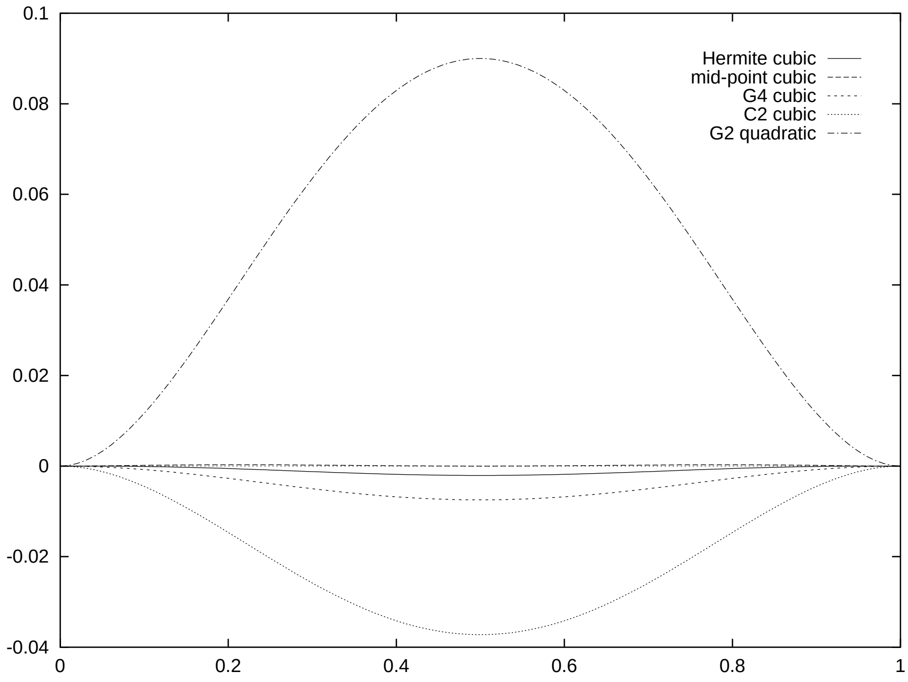

Figure 4. The polynomial $f_{\mathbf{r}}(\mathbf{q}(t))$ when w=0.8.

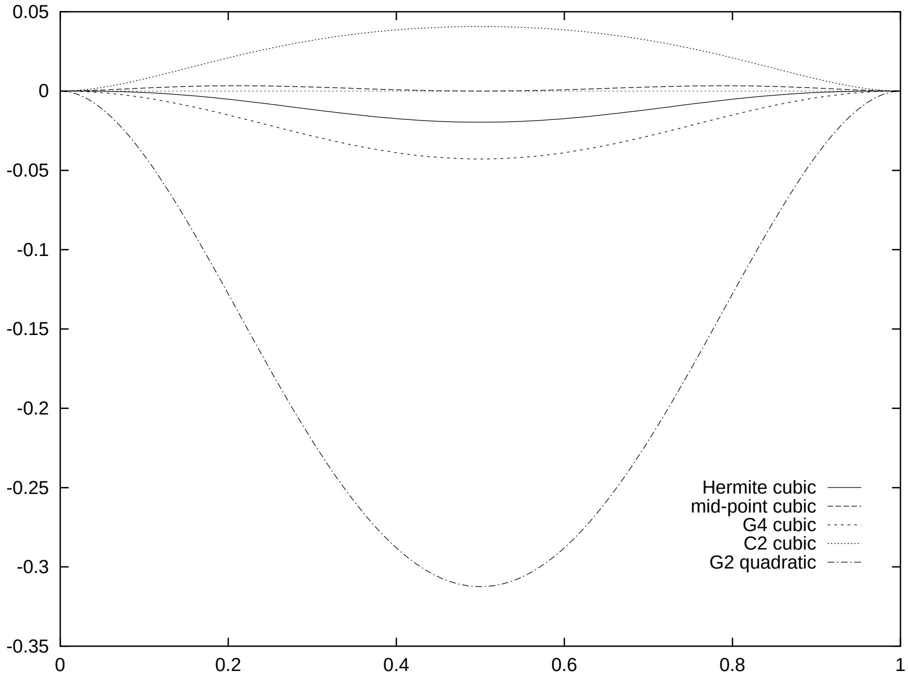

Figure 5. The polynomial $f_{\mathbf{r}}(\mathbf{q}(t))$ when w=1.5.

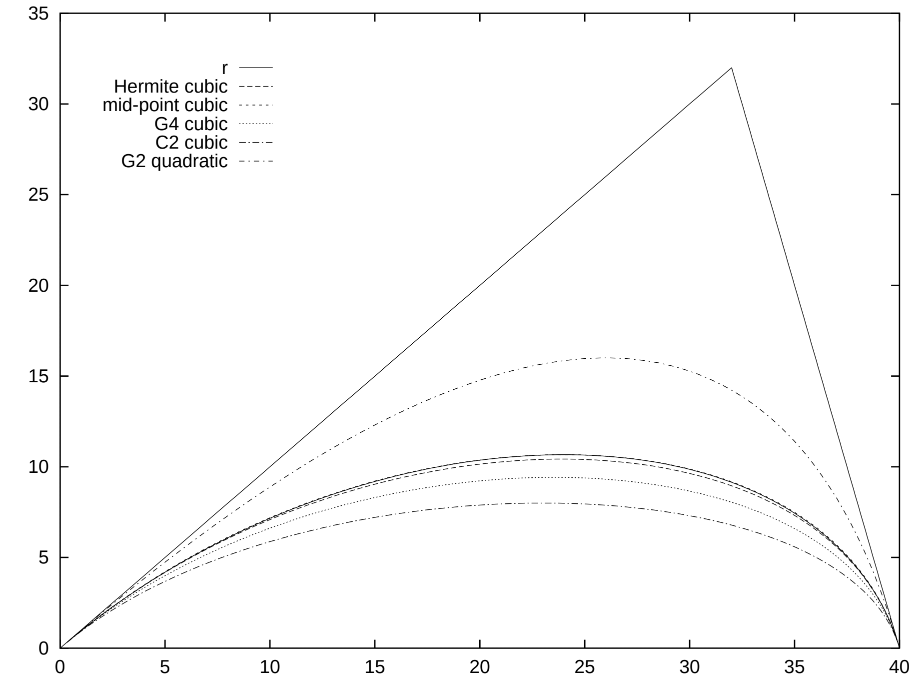

Figure 6. The curve $\mathbf{r}$ and the approximations when w=0.5.

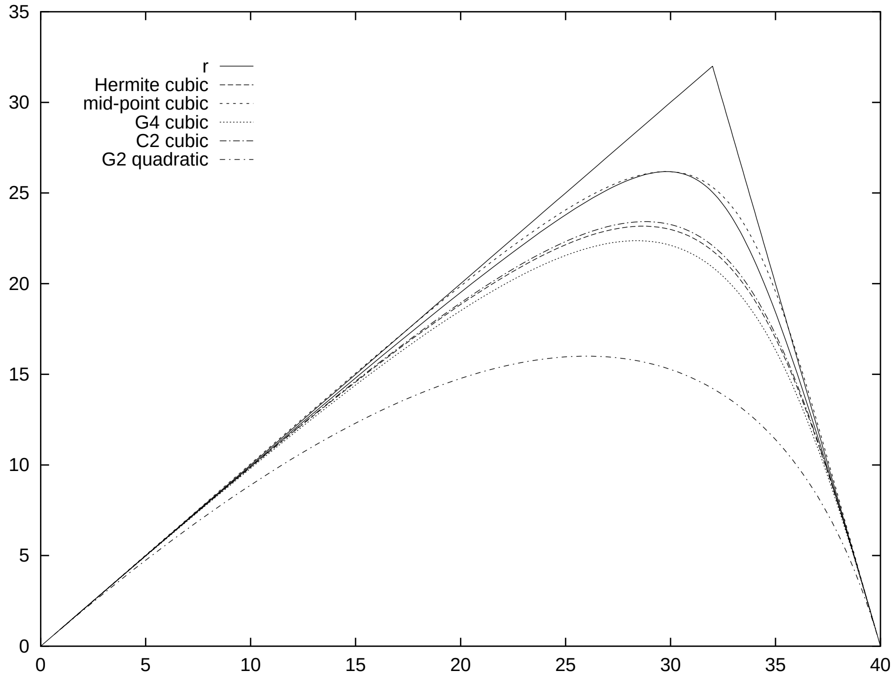

Figure 7. The curve **r** and the approximations when w=4.5.

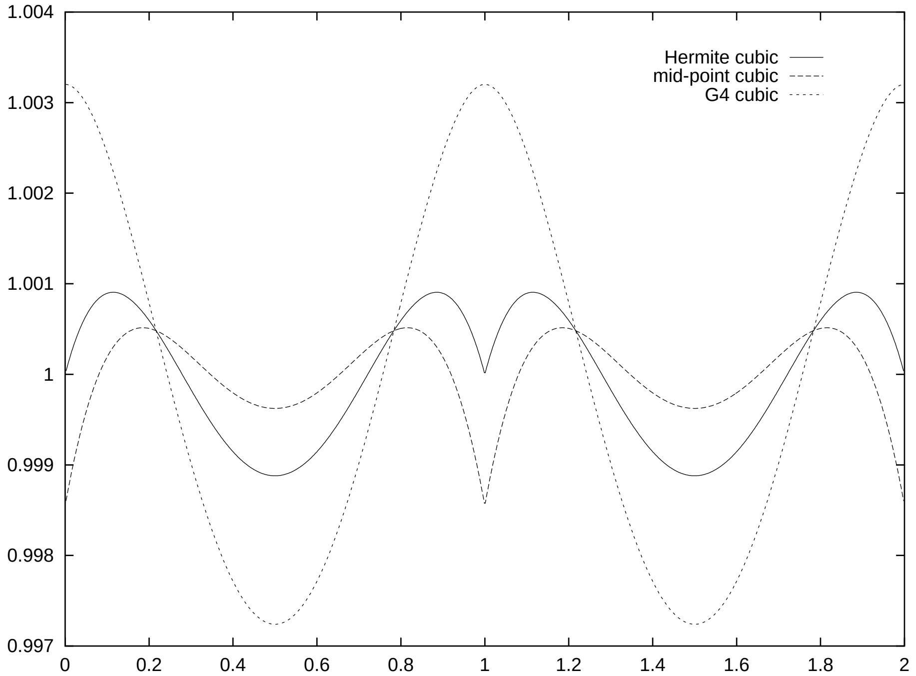

Figure 8. The curvature of the three sixth order approximations of two adjacent 45 degree circular arcs.