High order approximation of conic sections by quadratic splines

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

December 1993. Revised, October 1994

Abstract. Given a segment of a conic section in the form of a rational Bézier curve, a quadratic spline approximation is constructed and an *explicit* error bound is derived. The convergence order of the error bound is shown to be $O(h^4)$ which is optimal, and the spline curve is both C^1 and G^2. The approximation method is very efficient as it is based on local Hermite interpolation and subdivision. The approximation method and error bound are also applied to an important subclass of rational biquadratic surfaces which includes the sphere, ellipsoid, torus, cone and cylinder.

Keywords. approximation, conic sections, quadratic splines

§1. Introduction

The approximate conversion of rational splines to polynomial splines is an important requirement in computer-aided design. It is often necessary to transfer data from one design system to another even when they use different representations. Also a number of algorithms are difficult to generalise from polynomial splines to rational splines, for example lofting and blending, because of the necessity of positive weights. Evaluation and intersection algorithms are less efficient for rational splines. So even though NURBS have been described as the 'geometry standard' for curve and surface modelling, it is nevertheless worthwhile investigating how well one can approximate conics and quadrics by polynomial splines.

A number of papers have been written on the approximation of rational curves and surfaces by non-rational ones [1], [13], [14], [17], [20]. Though all of these have some strength of their own, none of them provide an *error bound* having *optimal order of convergence*. Bardis & Patrikalakis [1] point out the lack of bounds for rationals, as do Filip, Magedson, and Markot [8]. Sederberg and Kakimoto [20] have made some progress by providing an error bound for their 'moving control point' approximation but this is unfortunately not optimal.

Without an optimal error bound, the approximation will inevitably contain more data than what is actually necessary for the approximation to be within the given tolerance. In the worse situation where no error bound is available, one would simply have to guess the required number of subdivisions, if employing a Hermite approximation, or the number of interpolation points, if using a global spline approximation. The reason for the difficulty with rationals is that expressions for derivatives are complicated by the denominator. Yet error bounds for spline approximation normally require bounds on some derivatives of the curve or surface in question.

In classical spline approximation, the optimal approximation order when approximating a curve $\mathbf{r}$ by a spline curve $\mathbf{q}$ with degree n is $O(h^{n+1})$. That is to say, if h is the maximum length of parameter interval, and the correct approximation method is chosen, there exists some constant K for which

$$\max_t |\mathbf{q}(t) - \mathbf{r}(t)| \leq K h^{n+1}. \tag{1}$$

The power of h cannot be increased. Put in simple terms, each time h is halved, the new error is roughly $2^{-(n+1)}$ times the previous one. For example, both C^1 cubic Hermite interpolation and, provided care is taken near the end points of the parameter domain, C^2 cubic spline interpolation are $O(h^4)$ as explained by de Boor [2]. In the cubic Hermite case, K depends on the fourth derivative of $\mathbf{r}$.

More recently, investigations into so called *parametric* or *geometric* approximation suggest that if $\mathbf{r}$ is a planar curve, then an approximation order of $O(h^{2n})$ is both attainable and optimal, at least under some restrictions on $\mathbf{r}$ such as convexity. By considering the approximation of a circular segment in a neighbourhood of a point, Lyche & Morken [15] have derived a local polynomial expansion of order $O(h^{2n})$ when n is odd. Very recently, the author [11] has constructed global Hermite interpolations for conic sections with $O(h^{2n})$ convergence. In parametric approximation one exploits the spare degrees of freedom which become available when one weakens the definition of the error to be a true measure of distance between the curves. For example, one might try to bound the distance of $\mathbf{q}$ from $\mathbf{r}$:

$$\max_t \min_s |\mathbf{q}(s) - \mathbf{r}(t)| \leq K h^{2n}. \tag{2}$$

Explicit error bounds for parametric approximation have not been derived at all due to the nonlinearity involved. In the paper by de Boor, Höllig, Sabin [3], it was shown that parametric cubic Hermite approximation of planar curves, when it is possible, is $O(h^6)$. But no explicit expression for K is available, although it is known to depend on the sixth derivative of $\mathbf{r}$. The same problem is true of the $O(h^4)$ approximation due to Schaback [19].

The approximation of circular arcs by cubic Bézier segments has been analysed by Dokken et al. [5]. Their approach is to represent the unit circle in its implicit form $x^2(t) + y^2(t) = 1$. One then argues that if $x^2(t) + y^2(t) - 1$ is small then the approximation is good. By this method they obtained an $O(h^6)$ error, smaller than that of de Boor, Höllig, Sabin in this special case. The implicit error bound leads to an explicit one when the approximation is close. Using a similar method, Mørken [16] has since constructed various fourth order approximations, i.e. $O(h^4)$ when approximating a circle by quadratic segments.

So where does this leave the spline approximation of a rational polynomial curve $\mathbf{r}$? Whether one carries out a classical or a parametric approximation, we do not have any error bounds in the usual sense. Though bounds have been derived for first derivatives [9] [10], it is doubtful whether these can be generalised to higher derivatives in a useful way.

Somehow, one feels that, rather than try to bound derivatives, it ought to be possible to exploit the special form of a rational polynomial in order to construct a good error bound with respect to a particular approximation method. Knowing already what the approximation order is for general curves, the goal should be to construct an error bound having the *same* *order of convergence*. The bound does not need to be in the form Kh^k. For example, in the parametric case, if we can show that

$$\max_t \min_s |\mathbf{q}(s) - \mathbf{r}(t)| \leq \epsilon(h),$$

and that there exists K for which $\epsilon(h) \leq Kh^{2n}$, then we will have an excellent method for approximating rationals. One should still try to find an $\epsilon(h)$ which is as small as possible among those which have optimal convergence order.

Sederberg and Kakimoto [20] give an upper bound on the error $\mathbf{q}(t) - \mathbf{r}(t)$ of their 'moving control point' polynomial approximation $\mathbf{q}$. This method is good in that it applies to any degree, but in the case when $\mathbf{q}$ is quadratic, the error is only $O(h^2)$ which is neither optimal in the sense of (1) nor of (2). This is borne out by the slow convergence shown in the numerical results at the end of that paper.

In the present paper we attack the problem in a different way. A restricted yet important class of rational polynomials, namely conic sections and surfaces formed from them, for which an explicit optimal error bound *can* be constructed is studied. We approximate a rational quadratic Bézier curve $\mathbf{r}$ by geometric quadratic Hermite interpolation and derive an error bound having order of convergence $O(h^4)$. This is *optimal* since we know from the work of Degen [4] that this type of approximation is $O(h^4)$, agreeing with (2). Moreover, the numerical examples indicate that the error bound is sharp in the limit, i.e. the difference between the error and the error bound becomes negligible in relation to the size of the error. Thus the algorithm we present here is guaranteed to produce no more data than is absolutely necessary when approximating conics by quadratic splines.

Since $\mathbf{r}$ is a rational Bézier curve, the approximation is very simple. One starts by approximating $\mathbf{r}$ by that quadratic Bézier curve $\mathbf{q}_0$ having the same three control points. If the error is small enough one stops here. Otherwise $\mathbf{r}$ is subdivided at the mid point, each subcurve is normalised and then approximated by a corresponding Bézier curve in the same way as before, resulting in a spline approximation $\mathbf{q}_1$. The process continues — subdividing and normalising — until the error between $\mathbf{q}_r$ and $\mathbf{r}$ is small enough. By subdividing in this non-linear way the spline approximation turns out to be both C^1 and G^2. Thus this method yields both optimal convergence order and optimal smoothness. This also means that $\mathbf{q}_r$ is a special case of the G^2 spline developed by Schaback [19]. Normally, to construct a G^2 quadratic spline through an arbitrary sequence of points requires the solution of a non-linear system; a shooting technique is proposed in [19].

An upper bound on the maximum distance of $\mathbf{q}_r$ from $\mathbf{r}$ in each interval is found by using an explicit parametrisation $s(t)$ for which the error between $\mathbf{q}_r(s(t))$ and $\mathbf{r}(t)$ can easily be computed; see Section 2. To be precise, a number ϵ_r is found such that

$$\max_t \min_s |\mathbf{q}_r(s) - \mathbf{r}(t)| \leq \epsilon_r,$$

and ϵ_r is $O(2^{-4r})$ where r is the level of subdivision.

Notice that $O(2^{-kr})$ implies $O(h^k)$. Indeed, no matter where the subdivision points are chosen, there are 2^r parameter intervals after r levels of subdivision. So let $h_{i,r}$ be the length of the i-th interval and $h_r = \max_i h_{i,r}$. Then

$$h_0 = h_{1,0} = \sum_{i=1}^{2^r} h_{i,r} \leq \sum_{i=1}^{2^r} h_r = 2^r h_r.$$

If the approximation error, δ_r say, is $O(2^{-kr})$ then there exists a constant $K > 0$ for which $\delta_r \leq 2^{-kr}K$. Therefore

$$\delta_r/h_r^k \leq 2^{kr}\delta_r/h_0^k \leq K/h_0^k$$

and this means that δ_r is $O(h_r^k)$.

Thus it follows that ϵ_r above is also $O(h^4)$, where h is the maximum length of parameter intervals of $\mathbf{r}$, with respect to the original parameterisation. The subdivision scheme is described formally in Section 3 and it is proved in Section 4 that ϵ_r is $O(2^{-4r})$ as $r \to \infty$.

In Section 5 the question of continuity is addressed. It is shown that, due to the construction, the spline approximation has both C^1 and G^2 continuity, even though it only has contact of order 1 with the rational curve at the subdivision points. This must be a special characteristic of conic sections. One would normally expect to achieve at most G^1 continuity when approximating locally with quadratic splines — there are only six degrees of freedom in a quadratic Bézier segment. The extra order of continuity is also a consequence of the way the subdivision scheme is chosen; alternating between subdivision at mid points and normalising. The scheme exploits the natural symmetry of the rational curve in normal form. In the special case when the rational curve is a circular arc, the subdivision scheme corresponds precisely to uniform subdivision with respect to angle or arc length.

In Section 6 it is explained how, with very little extra effort, the approximation method can also be applied to a large class of rational biquadratic surfaces and an optimal error bound is again derived. One obtains a G^2 biquadratic spline approximation which is fourth order accurate in each parameter direction. Numerical examples are presented in Section 7.

§2. The error bound

We will consider the approximation of the rational quadratic Bézier curve

$$\mathbf{r}(t) = \frac{B_0(t)\mathbf{p}_0 + B_1(t)w\mathbf{p}_1 + B_2(t)\mathbf{p}_2}{B_0(t) + B_1(t)w + B_2(t)} \tag{3}$$

where $\mathbf{p}_0,\ \mathbf{p}_1,\ \mathbf{p}_2 \in \mathbb{R}^2$ are the *control points*, $w \in \mathbb{R}$ is the *weight* associated with $\mathbf{p}_1$, assumed positive, $B_0(t) = (1-t)^2$, $B_1(t) = 2(1-t)t$, $B_2(t) = t^2$, are the *Bernstein basis functions*, and t is in the range [0,1]. The most general form of a rational curve of degree two is

$$\frac{B_0(t)w_0\mathbf{p}_0 + B_1(t)w_1\mathbf{p}_1 + B_2(t)w_2\mathbf{p}_2}{B_0w_0(t) + B_1(t)w_1 + B_2(t)w_2}$$

for arbitrary $w_0,\ w_1,\ w_2 > 0$ but the so-called *normal form* (3) can always be arranged by a scaling and reparametrisation; see Piegl & Tiller [18]. It is shown in Faux & Pratt [7] that in normal form the size of w determines the type of conic section $\mathbf{r}$ represents. $\mathbf{r}$ is an ellipse when $w < 1$, a parabola when $w = 1$ and a hyperbola when $w > 1$; see Figure 1. The quantity $a = w - 1$ will play an important role in the analysis which follows.

We shall be concerned with the maximum error $|\mathbf{q}(s) - \mathbf{r}(t)|$ where $\mathbf{q}$ is the Bézier curve

$$\mathbf{q}(s) = B_0(s)\mathbf{p}_0 + B_1(s)\mathbf{p}_1 + B_2(s)\mathbf{p}_2$$

with $s \in [0,1]$. Curves $\mathbf{r}$ with $w < 1$ and $\mathbf{q}$ are shown in Figure 2. The points $\mathbf{r}(1/2)$ and $\mathbf{q}(1/2)$, known as the *shoulder points* of $\mathbf{r}$ and $\mathbf{q}$ respectively, both lie on the straight line

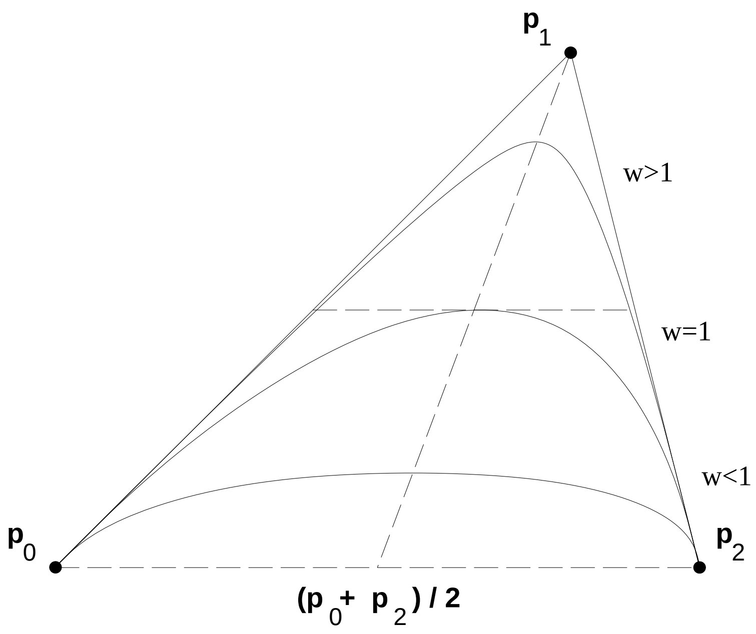

Figure 1. $\mathbf{r}$ in the cases $0<w<1$, $w=1$ and $w>1$.

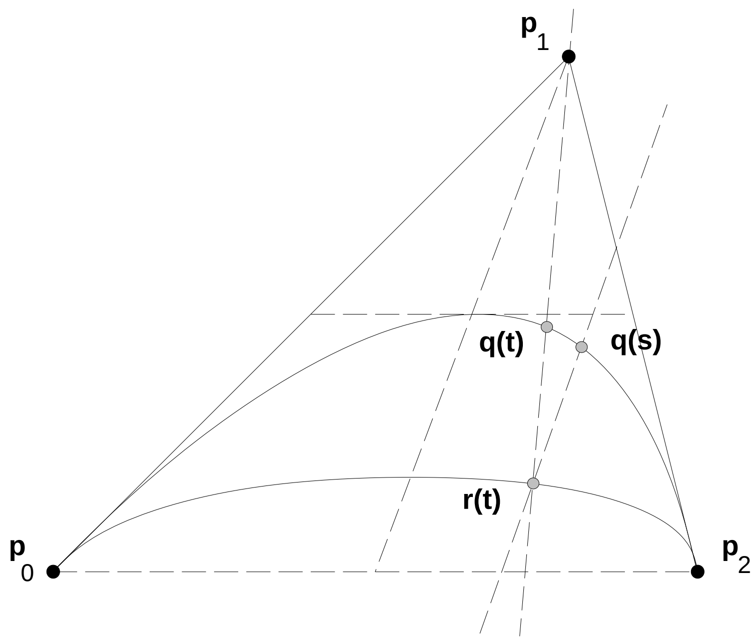

Figure 2. The curves $\mathbf{r}$ and $\mathbf{q}$.

between $(\mathbf{p}_0+\mathbf{p}_2)/2$ and $\mathbf{p}_1$. One might expect that $|\mathbf{q}(t)-\mathbf{r}(t)|$ would achieve its maximum when $t = 1/2$ but a calculation reveals that that this is in general not the case. Thus we dismiss the usual type of approximation $\mathbf{q}(t) \approx \mathbf{r}(t)$ and turn instead to a parametric approximation. Indeed we study a reparametrisation $s(t)$ for which $|\mathbf{q}(s(t)) - \mathbf{r}(t)|$ does achieve its maximum at $t = s = 1/2$. We define $s(t)$ by demanding that the vector $\mathbf{q}(s(t)) - \mathbf{r}(t)$ is parallel with $\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2$.

Proposition 2.1. *Let* $s = t(1 + a(1 - t))/(1 + 2a(1 - t)t)$ *where* $a = w - 1 > -1$. *Then*

$$\mathbf{q}(s) - \mathbf{r}(t) = \frac{a(2 + a)(1 - t)^2 t^2}{(1 + 2a(1 - t)t)^2}(\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2).$$

Proof. Since $s = t(1 + a(1 - t))/(1 + aB_1(t))$, and $1 - s = (1 - t)(1 + at)/(1 + aB_1(t))$, it follows that

$$\begin{aligned} B_0(s) =& B_0(t)(1 + at)^2/(1 + aB_1(t))^2, \\ B_1(s) =& B_1(t)(1 + at)(1 + a(1 - t))/(1 + aB_1(t))^2, \\ B_2(s) =& B_2(t)(1 + a(1 - t))^2/(1 + aB_1(t))^2. \end{aligned}$$

Now

$$\mathbf{r}(t) = \frac{B_0(t)\mathbf{p}_0 + (1 + a)B_1(t)\mathbf{p}_1 + B_2(t)\mathbf{p}_2}{1 + aB_1(t)},$$

so that

$$\begin{aligned} &(1 + aB_1(t))^2(\mathbf{q}(s) - \mathbf{r}(t)) \\ &\quad= \{(1 + at)^2 - (1 + aB_1(t))\}B_0(t)\mathbf{p}_0 \\ &\qquad + \{(1 + at)(1 + a(1 - t)) - (1 + a)(1 + aB_1(t))\}B_1(t)\mathbf{p}_1 \\ &\qquad + \{(1 + a(1 - t))^2 - (1 + aB_1(t))\}B_2(t)\mathbf{p}_2 \\ &\quad= a(2 + a)t^2 B_0(t)\mathbf{p}_0 - a(2 + a)(1 - t)tB_1(t)\mathbf{p}_1 + a(2 + a)(1 - t)^2 B_2(t)\mathbf{p}_2 \\ &\quad= a(2 + a)(1 - t)^2 t^2(\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2), \end{aligned}$$

as claimed. ◁

The validity of the reparametrisation can be verified by noticing that both s and $1 - s$ are positive whenever $0 < t < 1$ because $a > -1$. Also the denominator is always positive. The first derivative is found to be

$$s'(t) = (1 + a(1 - 2t(1 - t)))/(1 + 2a(1 - t)t)^2,$$

and so $s'(t) > 0$. Note that the vectors $\mathbf{q}(s(t)) - \mathbf{r}(t)$ and $\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2$ have the same direction when $a > 0$ and the opposite when $a < 0$. The magnitude of $\mathbf{q}(s(t)) - \mathbf{r}(t)$ is bounded in the following corollary.

Corollary 2.2. *With* $s = s(t)$ *as defined in Proposition 2.1,*

$$|\mathbf{q}(s) - \mathbf{r}(t)| \leq |\mathbf{q}(1/2) - \mathbf{r}(1/2)| = \frac{|a|}{4(2 + a)}|\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2|$$

for all $t \in [0, 1]$.

Proof. From Proposition 2.1 we find that

$$|\mathbf{q}(s) - \mathbf{r}(t)| = \frac{|a|(2 + a)B_1^2(t)}{4(1 + aB_1(t))^2}|\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2|.$$

If $\phi(t) = B_1(t)/(1 + aB_1(t))$ then $\phi'(t) = B_1'(t)/(1 + aB_1(t))^2$ and so, since $B_1'(1/2) = 0$, ϕ takes its maximum in $[0, 1]$ at $t = 1/2$. Therefore $|\mathbf{q}(s) - \mathbf{r}(t)|$ takes its maximum value when $t = s = 1/2$. Since $B_1(1/2) = 1/2$ and $\phi(1/2) = 1/(2 + a)$ the corollary is proven. ◁

Remark. An alternative way of deriving the error bound is to affinely map $\mathbf{r}$ into either a circular arc or an equilateral hyperbola $\mathbf{x}$. $\mathbf{x}$ can be parametrized in terms of either the trigonometric or hyperbolic functions repsectively: $\mathbf{x}(t) = \rho(c(t), s(t))$. In these cases the error bound in the corollary is the error between $\mathbf{x}$ and its parabolic approximation along the x axis and is found after some algebra to be

$$\epsilon = \frac{\rho}{2}\left(c(h) + \frac{1}{c(h)} - 2\right) = \frac{\rho}{8}h^4 + O(h^6),$$

where h is half the length of the parameter interval spanning $\mathbf{x}$. Since this error bound is a ratio, it is invariant under the affine mapping and so it is also valid for $\mathbf{r}$. For example, when $\mathbf{r}$ is an elliptic arc, $\mathbf{x}$ becomes a circular arc. If θ is the angle subtended by the circular arc, $h = 2\theta$ and we find that

$$\epsilon \approx \rho\frac{\theta^4}{128},$$

showing that the error is $O(\theta^4)$.

§3. Approximation by subdivision

Corollary 2.2 gives an upper bound on the error when approximating $\mathbf{r}$ by a single Bézier segment $\mathbf{q}$ which from now one will be referred to as $\mathbf{q}_0$. Indeed it was demonstrated that $d(\mathbf{q}_0, \mathbf{r}) \leq \epsilon_0$ where

$$d(\mathbf{q}_0, \mathbf{r}) = \max_{t\in[0,1]} \min_{s\in[0,1]} |\mathbf{q}_0(s) - \mathbf{r}(t)|$$

is the maximum distance of $\mathbf{q}_0$ from $\mathbf{r}$ and

$$\epsilon_0 = |a||\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2|/4(2 + a).$$

One can use this bound to obtain an arbitrarily close spline approximation to $\mathbf{r}$ in the form of a sequence of Bézier segments by recursive binary subdivision. By subdividing $\mathbf{r}$ at the right points we will see that the segments join with order of continuity C^1 and G^2. Indeed the subdivision scheme consists of alternating between subdividing at mid points and normalising the new segments. The term G^2 refers to geometric continuity of order 2. In this paper, a parametric curve will be said to be G^2 if there exists a reparametrisation for which it is C^2.

$\mathbf{r}$ is subdivided by the rational de Casteljau algorithm; see Farin [6]. Letting $\mathbf{r}_1$ be the subcurve $\mathbf{r}|_{0\leq t\leq 1/2}$ and $\mathbf{r}_2$ be the subcurve $\mathbf{r}|_{1/2\leq t\leq 1}$ one finds

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

and

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

for $t \in [0, 1]$ where $v = (1 + w)/2$, and

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

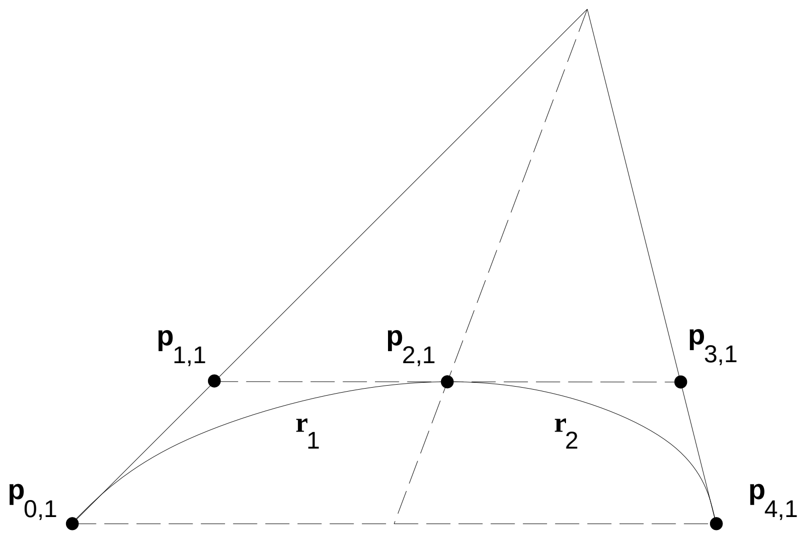

Figure 3. The first subdivision of $\mathbf{r}$ (with $w<1$).

see Figure 3.

Reparameterising $\mathbf{r}_1$ and $\mathbf{r}_2$ to put them in normal form, we obtain

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

and

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

where $w_1 = \sqrt{v} = \sqrt{(1+w)/2}$. If we approximate $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$ by

$$\mathbf{q}_{1,1}(s) = B_0(s)\mathbf{p}_{0,1} + B_1(s)\mathbf{p}_{1,1} + B_2(s)\mathbf{p}_{2,1}$$

and

$$\mathbf{q}_{2,1}(s) = B_0(s)\mathbf{p}_{2,1} + B_1(s)\mathbf{p}_{3,1} + B_2(s)\mathbf{p}_{4,1}$$

respectively, we can apply Corollary 2.2 again and obtain $d(\mathbf{q}_{1,1}, \hat{\mathbf{r}}_1) \le \epsilon_{1,1}$, $d(\mathbf{q}_{2,1}, \hat{\mathbf{r}}_2) \le \epsilon_{2,1}$, where

$$\epsilon_{1,1} = |a_1||\mathbf{p}_{0,1} - 2\mathbf{p}_{1,1} + \mathbf{p}_{2,1}|/4(2+a_1), \qquad \epsilon_{2,1} = |a_1||\mathbf{p}_{2,1} - 2\mathbf{p}_{3,1} + \mathbf{p}_{4,1}|/4(2+a_1),$$

and $a_1 = w_1 - 1$. It is a consequence of the above bounds that the approximation $\mathbf{q}_1$ consisting of the two segments $\mathbf{q}_{1,1}$ and $\mathbf{q}_{2,1}$ is such that $d(\mathbf{q}_1, \mathbf{r}) \le \epsilon_1 = \max(\epsilon_{1,1}, \epsilon_{2,1})$.

Subdivision algorithm. By continuing to subdivide at the midpoint of each new normalised segment we obtain the subdivision scheme:

(i) Set $\mathbf{p}_{i,0} = \mathbf{p}_i$ for $i = 0, 1, 2$ and $w_0 = w$.

(ii) For $r = 1, 2, \ldots$

let, for $i = 0, \ldots, 2^{r-1} - 1$,

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

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

The r-th approximation $\mathbf{q}_r(s)$, $s \in [0, 1]$ to $\mathbf{r}$ is the piecewise quadratic

$$\mathbf{q}_r(s) = B_0(\xi)\mathbf{p}_{2i,r} + B_1(\xi)\mathbf{p}_{2i+1,r} + B_2(\xi)\mathbf{p}_{2i+2,r}$$

where $\xi = 2^r s - i$, for $s \in [i/2^r, (i+1)/2^r]$, $i = 0, \dots, 2^r - 1$.

§4. Order of approximation

In order to study the convergence of the error bound, we define $a_r = w_r - 1$ and

$$\epsilon_r = \frac{|a_r|}{4(2 + a_r)} \max_{i=0,\dots,2^r-1} |\mathbf{p}_{2i,r} - 2\mathbf{p}_{2i+1,r} + \mathbf{p}_{2i+2,r}|. \tag{5}$$

Due to the construction of $\mathbf{q}_r$ and Corollary 2.2, we have that $d(\mathbf{q}_r, \mathbf{r}) \leq \epsilon_r$. In the following theorem it is shown that $\epsilon_r \to 0$ as $r \to \infty$ and the convergence is fourth order. This of course implies that $d(\mathbf{q}_r, \mathbf{r})$ also has fourth order convergence but this follows from a theorem in Degen [4] since $\mathbf{q}_r$ is a quadratic Hermite interpolation in each segment. The theorem actually looks a lot more complicated than it really is. Essentially the two significant components $|a_r|$ and $\max_{i=0,\dots,2^r-1} |\mathbf{p}_{2i,r} - 2\mathbf{p}_{2i+1,r} + \mathbf{p}_{2i+2,r}|$ are both $O(2^{-2r})$, i.e. $O(h^2)$.

The quantity $|a_r|$ is in a sense a measure of the "rationality" of each subcurve of $\mathbf{r}$ after r subdivision levels. Indeed, $a_r = w_r - 1$ and if a_r were 0 (equivalently a were 0), $\mathbf{r}$ would be a parabola. Thus it is not surprising that if $|a_r|$ is $O(h^2)$, as shown in the theorem, that $\mathbf{q}_r$ approaches $\mathbf{r}$ as r increases — $\mathbf{q}_r$ can be regarded as a rational curve with $a_r = 0$. The first part of the theorem shows in fact that the "rationality" of each curve segment gets smaller at a rate of $O(h^2)$. This turns out to be quite easy to prove because a_r can be solved explicitly.

Meanwhile the quantity $|\mathbf{p}_{2i,r} - 2\mathbf{p}_{2i+1,r} + \mathbf{p}_{2i+2,r}|$, is clearly a second order difference and one would expect this to be $O(h^2)$. Now, had $\mathbf{r}$ been non-rational ($a = a_r = 0$ for all r), it would have been almost trivial to prove this. One could either use the fact that $2(\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2)$ were precisely equal to the constant second derivative $\mathbf{r}''(t)$ or from the subdivision scheme (4), one would obtain

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

But since $a_r \neq 0$, one finds instead equation (6) below. Even so it is relatively straightforward to demonstrate that α_r is $O(2^{-(2-\delta)r})$ for any small δ. Most of the complications arise when achieving $O(2^{-2r})$.

Theorem 4.1. *The error bound ϵ_r is $O(2^{-4r})$ as $r \to \infty$.*

Proof. The proof is in two parts. We show that both (i) $|a_r|$ is $O(2^{-2r})$ and (ii) $\max_i |\mathbf{p}_{2i,r} - 2\mathbf{p}_{2i+1,r} + \mathbf{p}_{2i+2,r}|/(2 + a_r)$ is $O(2^{-2r})$ as $r \to \infty$.

(i) The recursive equation for w_r resembles the half angle formulas for cos and cosh, namely

$$\cos(x/2) = \sqrt{(\cos x + 1)/2}, \qquad \text{and} \qquad \cosh(x/2) = \sqrt{(\cosh x + 1)/2}.$$

From this observation we can solve w_r explicitly. If $w_0 < 1$ we find

$$w_r = \cos(2^{-r} \cos^{-1}(w_0)),$$

and when $w_0 > 1$,

$$w_r = \cosh(2^{-r}\cosh^{-1}(w_0)),$$

Now, taking the case $w_0 < 1$, if we let $\theta = \cos^{-1}(w_0)$, we find by expanding cos in its power series that

$$2^{2r}a_r = 2^{2r}(w_r - 1) = -\frac{\theta^2}{2!} + \frac{\theta^4}{4!\ 2^{2r}} - \frac{\theta^6}{6!\ 2^{4r}} \cdots.$$

Therefore $2^{2r}a_r$ is bounded and indeed

$$2^{2r}a_r \to -(\cos^{-1}(w_0))^2/2.$$

A similar argument shows that $2^{2r}a_r$ is also bounded when $w_0 > 1$ and then

$$2^{2r}a_r \to (\cosh^{-1}(w_0))^2/2.$$

Thus $|a_r|$ is $O(2^{-2r})$ as claimed.

(ii) To improve clarity, set

$$\begin{aligned}\alpha_r &= \max_{i=0,\ldots,2^r-1} |\mathbf{p}_{2i-2,r} - 2\mathbf{p}_{2i-1,r} + \mathbf{p}_{2i,r}|/(2+a_r)\\ \beta_r &= \max_{i=0,\ldots,2^{r+1}-1} |\mathbf{p}_{i+1,r} - \mathbf{p}_{i,r}|/(2+a_r).\end{aligned}$$

The task is to show that α_r is $O(2^{-2r})$. From the subdivision scheme (4) we find that after some manipulation,

$$\mathbf{p}_{4i,r} - 2\mathbf{p}_{4i+1,r} + \mathbf{p}_{4i+2,r} = \frac{\mathbf{p}_{2i,r-1} - 2\mathbf{p}_{2i+1,r-1} + \mathbf{p}_{2i+2,r-1}}{2(2+a_{r-1})} + \frac{a_{r-1}(\mathbf{p}_{2i,r-1} - \mathbf{p}_{2i+1,r-1})}{(2+a_{r-1})},$$

and

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

Taking maximums over each side and dividing by $(2 + a_r)$, we then have

$$\alpha_r \le \frac{\alpha_{r-1}}{2(2+a_r)} + \frac{|a_{r-1}|\beta_{r-1}}{(2+a_r)}. \tag{6}$$

It turns out that $|a_{r-1}|\beta_{r-1}$ becomes negligible relative to α_{r-1} in the limit. Further algebra reveals that

$$\mathbf{p}_{4i+1,r} - \mathbf{p}_{4i,r} = \frac{(1+a_{r-1})(\mathbf{p}_{2i+1,r-1} - \mathbf{p}_{2i,r-1})}{(2+a_{r-1})},$$

and

$$\mathbf{p}_{4i+2,r} - \mathbf{p}_{4i+1,r} = \frac{\mathbf{p}_{2i+2,r-1} - \mathbf{p}_{2i,r-1}}{2(2+a_{r-1})} = \frac{\mathbf{p}_{2i+2,r-1} - \mathbf{p}_{2i+1,r-1}}{2(2+a_{r-1})} + \frac{\mathbf{p}_{2i+1,r-1} - \mathbf{p}_{2i,r-1}}{2(2+a_{r-1})}.$$

With the other two cases being symmetries of these, taking the maximum over i of each side and dividing by $(2 + a_r)$ implies

$$\beta_r \leq \frac{1 + |a_{r-1}|}{2 + a_r}\beta_{r-1},$$

and in view of the fact that $a_r \to 0$ it is clear that β_r is $O(2^{-(1-\delta)r})$ for any $\delta > 0$. To obtain the sharper $O(2^{-r})$, let $\phi_r = 2^r\beta_r$ and take logs. Using the fact that $\log(x) \leq x - 1$ for $x > 0$, one finds

$$\begin{aligned}\log(\phi_r) \leq& \log\left(\frac{2 + 2|a_{r-1}|}{2 + a_r}\right) + \log(\phi_{r-1})\\ \leq& \frac{2|a_{r-1}| - a_r}{2 + a_r} + \log(\phi_{r-1})\\ \leq& 3|a_{r-1}| + \log(\phi_{r-1}),\end{aligned}$$

since $|a_r| < |a_{r-1}|$ follows from (4) and $2 + a_r > 1$. Further, because $|a_r| \leq 2^{-2r}K$, for some K,

$$\log(\phi_r) \leq 4K + \log(\phi_0),$$

i.e.

$$\phi_r \leq e^{4K}\phi_0,$$

which means

$$\beta_r \leq 2^{-r}e^{4K}\beta_0,$$

and therefore β_r is $O(2^{-r})$.

Now we attack (6) in a similar way. Combining the convergence of β_r with a_r we see that the product $|a_r|\beta_r$ must be $O(2^{-3r})$. In other words there exists a constant L for which $2^{3r}|a_r|\beta_r \leq L$. Therefore if we let $\gamma_r = 2^{2r}\alpha_r$, we obtain

$$\gamma_r \leq \frac{2}{(2 + a_r)}(\gamma_{r-1} + 2^{-r+2}L).$$

As previously the key to showing that γ_r is bounded is to take logs:

$$\begin{aligned}\log(\gamma_r) \leq& \log\left(\frac{2}{2 + a_r}\right) + \log(\gamma_{r-1} + 2^{-r+2}L)\\ \leq& \frac{|a_r|}{2 + a_r} + \log(\gamma_{r-1} + 2^{-r+2}L).\end{aligned}$$

Now we have $\log(x) \leq x - 1$ and moreover, since log is concave, $\log(a+b) \leq \log(a) + b\log'(a)$ for $a, b > 0$. Then $\log(a + b) \leq \log(a) + b \leq |\log(a)| + b$ when $a \geq 1$ and $\log(a + b) \leq a + b - 1 \leq b \leq |\log(a)| + b$ when $a < 1$. In either case, $|\log(a + b)| \leq |\log(a)| + b$. Applying this estimate, one finds

$$|\log(\gamma_r)| \leq |a_r| + |\log(\gamma_{r-1})| + 2^{-r+2}L,$$

and so

$$\log(\gamma_r) \leq |\log(\gamma_r)| \leq K/3 + |\log(\gamma_0)| + 4L.$$

Thus

$$\gamma_r \leq e^{K/3+|\log(\gamma_0)|+4L},$$

and therefore

$$\alpha_r \leq 2^{-2r} e^{K/3+|\log(\alpha_0)|+4L}.$$

Hence α_r is $O(2^{-2r})$ as claimed. $\triangleleft$

§5. Order of continuity

Since the r-th approximation $\mathbf{q}_r$ touches $\mathbf{r}$ tangentially at the points of subdivision $\mathbf{p}_{0,r},\ \mathbf{p}_{2,r},\ \mathbf{p}_{4,r},\ \ldots,\ \mathbf{p}_{2^{r+1},r}$, it is clear that $\mathbf{q}_r$ and $\mathbf{r}$ meet with order of contact 1 and that $\mathbf{q}_r$ is itself a G_1 curve. In the following theorem it is shown that the order of continuity of $\mathbf{q}_r$ is in fact both C^1 and G^2.

Recall that sufficient conditions for the two Bézier curves $B_0(u)\mathbf{a}_0 + B_1(u)\mathbf{a}_1 + B_2(u)\mathbf{a}_2$ and $B_0(u)\mathbf{b}_0 + B_1(u)\mathbf{b}_1 + B_2(u)\mathbf{b}_2$ to join with C^1 and G^2 continuity at $u = 1$ and $u = 0$ respectively are $\mathbf{a}_2 = \mathbf{b}_0$, $\mathbf{a}_1 - 2\mathbf{a}_2 + \mathbf{b}_1 = 0$, and $\mathbf{b}_2 - \mathbf{a}_0 = \lambda(\mathbf{b}_1 - \mathbf{a}_1)$ for some scalar λ. Thus when $r = 1$, it can be seen from Figure 3 that the two subcurves of $\mathbf{q}_1$ join with C^1 and G^2 continuity. Mathematically, from the definition (4) of the $\mathbf{p}_{i,1}$, one finds

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

and

$$\mathbf{p}_{4,1} - \mathbf{p}_{0,1} = (1 + w_0)(\mathbf{p}_{3,1} - \mathbf{p}_{1,1}).$$

In other words, the vectors $\mathbf{p}_{2,1} - \mathbf{p}_{1,1}$ and $\mathbf{p}_{3,1} - \mathbf{p}_{2,1}$ are equal (in both length and direction) while the vectors $\mathbf{p}_{3,1} - \mathbf{p}_{1,1}$ and $\mathbf{p}_{4,1} - \mathbf{p}_{0,1}$ are parallel (the ratio of their lengths depends on the weight w_0). By a similar argument, after the next subdivision $r = 2$, the first two of the four subcurves of $\mathbf{q}_2$ join with C^1 and G^2 continuity as do the last two. The order of continuity between the middle two subcurves on the other hand follows from the continuity of the two subcurves of $\mathbf{q}_1$.

Theorem 5.1. *The curve $\mathbf{q}_r$ is both C^1 and G^2.*

Proof. We prove, by induction on r, that for all r and all i,

$$\mathbf{p}_{2i+1,r} - 2\mathbf{p}_{2i+2,r} + \mathbf{p}_{2i+3,r} = 0, \tag{7}$$

and

$$\mathbf{p}_{2i+4,r} - \mathbf{p}_{2i,r} = (1 + w_{r-1})(\mathbf{p}_{2i+3,r} - \mathbf{p}_{2i+1,r}) = 2w_r^2(\mathbf{p}_{2i+3,r} - \mathbf{p}_{2i+1,r}). \tag{8}$$

Let $r \geq 2$ and assume that these identities hold for $r - 1$.
To prove (7), for each $i \in \{0, \ldots, 2^{r-1} - 1\}$ there are two cases. From the subdivision scheme (4) we find

$$\mathbf{p}_{4i+1,r} - 2\mathbf{p}_{4i+2,r} + \mathbf{p}_{4i+3,r} = 0$$

and

$$\mathbf{p}_{4i+3,r} - 2\mathbf{p}_{4i+4,r} + \mathbf{p}_{4i+5,r} = \frac{w_{r-1}}{(1+w_{r-1})}(\mathbf{p}_{2i+1,r-1} - 2\mathbf{p}_{2i+2,r-1} + \mathbf{p}_{2i+3,r-1}) = 0$$

by the induction hypothesis.
To prove (8) there are again two cases. In the first we find

$$\mathbf{p}_{4i+4,r} - \mathbf{p}_{4i,r} = (1+w_{r-1})(\mathbf{p}_{4i+3,r} - \mathbf{p}_{4i+1,r})$$

directly from the subdivision scheme. In the second, by the induction hypothesis, $\mathbf{p}_{2i+4,r-1} - \mathbf{p}_{2i,r-1} = (1+w_{r-2})(\mathbf{p}_{2i+3,r-1} - \mathbf{p}_{2i+1,r-1})$. This implies that

$$\mathbf{p}_{4i+6,r} - \mathbf{p}_{4i+2,r} = \frac{2w_{r-1} + (1+w_{r-2})}{2w_{r-1}}(\mathbf{p}_{4i+5,r} - \mathbf{p}_{4i+3,r}) = (1+w_{r-1})(\mathbf{p}_{4i+5,r} - \mathbf{p}_{4i+3,r}),$$

as required.

When $r = 1$,

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

and

$$\mathbf{p}_{4,1} - \mathbf{p}_{0,1} = (1+w_0)(\mathbf{p}_{3,1} - \mathbf{p}_{1,1}),$$

which completes the proof. ◁

The above theorem means that we can express the r-th approximation $\mathbf{q}_r$ as a uniform quadratic spline. Its control points are

$$\mathbf{p}_{0,r},\ \mathbf{p}_{1,r},\ \mathbf{p}_{3,r},\ \mathbf{p}_{5,r},\ \ldots,\ \mathbf{p}_{2^{r+1}-1,r},\ \mathbf{p}_{2^{r+1},r}$$

and its knot vector is $0, 0, 1/2^r, 2/2^r, \ldots, (2^r-1)/2^r, 1, 1$.

Note that the C^1 and G^2 continuity of the approximant depend critically on the fact that $\mathbf{r}$ is subdivided in *every* interval simultaneously. If one subdivided adaptively, the approximant would in general only be G^1.

Remark. Note also that the G^2 continuity can be demonstrated alternatively by affinely mapping, at each level of subdivision, each pair of subcurves into a circular arc or equilateral hyperbola. Since then the two approximating curves are symmetries of each other, they share the same curvature at their contact point. The G^2 continuity then follows because curvature is an affinely-invariant quantity.

§6. Approximation of surfaces

Using the same error bound and essentially the same subdivision scheme as developed for curves one can obtain a fourth order approximation and error bound for members of an important subclass of rational tensor-product biquadratic Bézier surfaces. In fact, we can construct a biquadratic spline approximation to the parametric surface

$$\mathbf{r}(u,v) = \frac{\sum_{i=0}^{2}\sum_{j=0}^{2} B_i(u)B_j(v)w_{ij}\mathbf{p}_{ij}}{\sum_{i=0}^{2}\sum_{j=0}^{2} B_i(u)B_j(v)w_{ij}}$$

which is both fourth order accurate in each parameter direction and C^1 and G^2 provided only that

$$w_{00}w_{ij} = w_{i0}w_{0j}. \tag{9}$$

No restrictions whatsoever are put on the control points $\mathbf{p}_{ij}$. A surface is said to be G^2 if it can be reparametrised so as to be C^2.

Without demanding condition (9), the bound in (12) would not be possible. Furthermore the high order of continuity (both C^1 and G^2) would be lost as equations (13) and (14) would no longer be valid. There does not appear to be any easy way of improving on this. With condition (9), $\mathbf{r}$ has the property that every isoparametric curve in u (a surface curve of the form $v = const$) is a rational quadratic curve with the same three weights, namely w_{00}, w_{10}, w_{20}, *irrespectively* of v. For if one fixes v to be some $\bar{v}$, one can write (after multiplying throughout by w_{00})

$$\mathbf{r}(u,\bar{v}) = \frac{\sum_{i=0}^{2} B_i(u)w_{i0}\mathbf{r}_i(\bar{v})}{\sum_{i=0}^{2} B_i(u)w_{i0}} \qquad \text{where} \qquad \mathbf{r}_i(\bar{v}) = \frac{\sum_{j=0}^{2} B_j(\bar{v})w_{0j}\mathbf{p}_{ij}}{\sum_{j=0}^{2} B_j(\bar{v})w_{0j}}.$$

The same is true, of course, of isoparametric curves in the v variable. There is an interesting analogy between (9) and the condition for a tensor-product Bézier surface to be *translational*, namely $\mathbf{p}_{00} + \mathbf{p}_{ij} = \mathbf{p}_{i0} + \mathbf{p}_{0j}$, as defined in [6].

Among surfaces satisfying (9) are those constructed by revolving any conic section (representable by a rational Bézier curve) in the xz plane about the z axis, through an angle of less than 180 degrees. For example a sphere is typically represented by 8 patches of this kind while the torus can be represented by 16 of them; see [18]. Due to the symmetry of the patches the approximation method will yield an approximant to the whole sphere or torus which is C^1 and G^2. One must make sure that, in each parameter direction, the number of subdivisions in each patch is constant. Patches of cylinders, cones, elliptic cones and ellipsoids can also be expressed in the form of $\mathbf{r}(u,v)$ satisfying (9). Note that in all of these particular cases there is in addition some kind of symmetry among the *control points* but this is not a requirement for the approximation method described here.

Given that the weights are in the form (9) it is straightforward to reparametrise $\mathbf{r}$ if necessary in such a way that

$$[w_{ij}] = \begin{bmatrix} 1 & w_2 & 1 \\ w_1 & w_1w_2 & w_1 \\ 1 & w_2 & 1 \end{bmatrix},$$

(the indices go from 0 to 2). We will now approximate $\mathbf{r}(u,v)$ by

$$\mathbf{q}(s,t) = \sum_{i=0}^{2}\sum_{j=0}^{2} B_i(s)B_j(t)\mathbf{p}_{ij}$$

where

$$s = u(1+a_1(1-u))/(1+a_1B_1(u)) \qquad \text{and} \qquad t = v(1+a_2(1-v))/(1+a_2B_1(v)),$$

and $a_1 = w_1 - 1$, $a_2 = w_2 - 1$. To compute an upper bound on

$$d(\mathbf{q},\mathbf{r}) = \max_{(u,v)\in\Omega}\min_{(s,t)\in\Omega} |\mathbf{q}(s,t) - \mathbf{r}(u,v)|, \qquad \text{with } \Omega = [0,1]\times[0,1],$$

we define the intermediate surface

$$\mathbf{g}(u,t) = \frac{\sum_{i=0}^{2}\sum_{j=0}^{2} B_i(u)B_j(t)w_{i0}\mathbf{p}_{ij}}{\sum_{i=0}^{2} B_i(u)w_{i0}}$$

which is rational in u but non-rational in t. Observe now that

$$\mathbf{q}(s,t) - \mathbf{g}(u,t) = \sum_{j=0}^{2} B_j(t)\left[\sum_{i=0}^{2} B_i(s)\mathbf{p}_{ij} - \frac{\sum_{i=0}^{2} B_i(u)w_{i0}\mathbf{p}_{ij}}{\sum_{i=0}^{2} B_i(u)w_{i0}}\right],$$

and, from (9),

$$\mathbf{g}(u,t) - \mathbf{r}(u,v) = \sum_{i=0}^{2} B_i(u)w_{i0}\left[\sum_{j=0}^{2} B_j(t)\mathbf{p}_{ij} - \frac{\sum_{j=0}^{2} B_j(v)w_{0j}\mathbf{p}_{ij}}{\sum_{j=0}^{2} B_j(v)w_{0j}}\right] \Big/ \sum_{i=0}^{2} B_i(u)w_{i0}.$$

Now the two expressions in the square brackets are completely analogous to the difference $\mathbf{q}(s) - \mathbf{r}(t)$ considered for curves in Corollary 2.2. Appealing to Corollary 2.2 and by the convex hull property of non-rational and rational Bézier curves respectively it follows that

$$|\mathbf{q}(s,t) - \mathbf{g}(u,t)| \leq \frac{|a_1|}{4(2+a_1)} \max_{j=0,1,2} |\mathbf{p}_{0j} - 2\mathbf{p}_{1j} + \mathbf{p}_{2j}| \tag{10}$$

and

$$|\mathbf{g}(u,t) - \mathbf{r}(u,v)| \leq \frac{|a_2|}{4(2+a_2)} \max_{i=0,1,2} |\mathbf{p}_{i0} - 2\mathbf{p}_{i1} + \mathbf{p}_{i2}|. \tag{11}$$

Hence

$$\begin{aligned} d(\mathbf{q},\mathbf{r}) \leq & |\mathbf{q}(s,t) - \mathbf{r}(u,v)| \\ \leq & \frac{|a_1|}{4(2+a_1)} \max_{j=0,1,2} |\mathbf{p}_{0j} - 2\mathbf{p}_{1j} + \mathbf{p}_{2j}| + \frac{|a_2|}{4(2+a_2)} \max_{i=0,1,2} |\mathbf{p}_{i0} - 2\mathbf{p}_{i1} + \mathbf{p}_{i2}|. \end{aligned} \tag{12}$$

It is clear from this that by recursively subdividing and normalising $\mathbf{r}$ in each parameter direction in an analogous way to (4), the convergence order of the bound is four in each direction. After each subdivision one can compute each of the bounds (10) and (11) separately. The greater of the two error bounds can then be used to determine the parameter direction in which to subdivide next. Note that for a surface such as a cylinder one would only need to subdivide in one direction.

The approximant will also be C^1 and G^2. Let us illustrate the proof of continuity by showing that after subdividing $\mathbf{r}(u, v)$ once at $u = 1/2$, the approximant $\mathbf{q}(s, t)$, consisting of the two patches $\mathbf{q}_1$ and $\mathbf{q}_2$, is C^1 and G^2 along the edge $s = u = 1/2$.

The goal then is to prove that the two patches

$$\mathbf{q}_1(s,t) = \sum_{i=0}^{2}\sum_{j=0}^{2} B_i(s)B_j(t)\mathbf{q}_{ij},$$
$$\mathbf{q}_2(s,t) = \sum_{i=0}^{2}\sum_{j=0}^{2} B_i(s)B_j(t)\mathbf{q}_{2+ij},$$

join with C^1 and G^2 continuity at $s = 1$ and $s = 0$ respectively, where

$$\begin{aligned}
\mathbf{q}_{0,j} =& \mathbf{p}_{0,j}, \\
\mathbf{q}_{1,j} =& (\mathbf{p}_{0,j} + w_1\mathbf{p}_{1,j})/(1 + w_1), \\
\mathbf{q}_{2,j} =& (\mathbf{p}_{0,j} + 2w_1\mathbf{p}_{1,j} + \mathbf{p}_{2,j})/2(1 + w_1), \\
\mathbf{q}_{3,j} =& (w_1\mathbf{p}_{1,j} + \mathbf{p}_{2,j})/(1 + w_1), \\
\mathbf{q}_{4,j} =& \mathbf{p}_{2,j}.
\end{aligned}$$

Similar to the curve case one finds that

$$\mathbf{q}_{1,j} - 2\mathbf{q}_{2,j} + \mathbf{q}_{3,j} = 0, \tag{13}$$

and

$$\mathbf{q}_{4,j} - \mathbf{q}_{0,j} = (1 + w_1)(\mathbf{q}_{3,j} - \mathbf{q}_{1,j}). \tag{14}$$

The C^1 continuity is a consequence of (13) and G^2 continuity follows from both (13) and (14) where the important point is that the factor $(1 + w_1)$ is independent of j. Indeed C^1 continuity is a consequence of the fact that the first derivatives with respect to s of $\mathbf{q}_1$ and $\mathbf{q}_2$ are equal at $s = 1$ and $s = 0$ respectively;

$$\frac{\partial}{\partial s}\mathbf{q}_2(0,t) = 2\sum_{j=0}^{2} B_j(t)(\mathbf{q}_{3,j} - \mathbf{q}_{2,j}) = 2\sum_{j=0}^{2} B_j(t)(\mathbf{q}_{2,j} - \mathbf{q}_{1,j}) = \frac{\partial}{\partial s}\mathbf{q}_1(1,t).$$

The derivatives $\partial/\partial t$ and $\partial^2/\partial s\partial t$ also agree since both patches are C^1 in t along the edge.

Now consider G^2 continuity. Following Gregory [12], it is sufficient to find a C^2 mapping $\varphi : \mathbb{R} \to \mathbb{R}$, defined in a neighbourhood of $s = 0$, with $\varphi(0) = 1$ and $\varphi'(0) > 0$ for which all partial derivatives in s up to order two of the patches $\mathbf{q}_1(\varphi(s), t)$ and $\mathbf{q}_2(s, t)$ are equal at $s = 0$. It is not necessary to look at derivatives involving t. This is due to the fact that the patches are C^2 along the adjoining edge and that the direction of the s variable in the parameter plane is transversal to the knot line; continuity of cross derivatives comes automatically from differentiation. Let $\varphi(s) = 1 + s - a_1 s^2$ where $a_1 = w_1 - 1$. By the chain rule one finds

$$\frac{\partial}{\partial s}\mathbf{q}_1(\varphi(s), t) = (1 - 2a_1 s)\frac{\partial}{\partial \varphi}\mathbf{q}_1(\varphi, t),$$

and

$$\frac{\partial^2}{\partial s^2}\mathbf{q}_1(\varphi(s), t) = -2a_1\frac{\partial}{\partial \varphi}\mathbf{q}_1(\varphi, t) + (1 - 2a_1 s)^2\frac{\partial^2}{\partial \varphi^2}\mathbf{q}_1(\varphi, t).$$

Then, from (13),

$$\frac{\partial}{\partial s}\mathbf{q}_2(0, t) = 2\sum_{j=0}^{2} B_j(t)(\mathbf{q}_{3,j} - \mathbf{q}_{2,j}) = 2\sum_{j=0}^{2} B_j(t)(\mathbf{q}_{2,j} - \mathbf{q}_{1,j}) = \frac{\partial}{\partial s}\mathbf{q}_1(\varphi(0), t),$$

and, applying both (13) and (14),

$$\begin{aligned}\frac{\partial^2}{\partial s^2}\mathbf{q}_2(0, t) =& 2\sum_{j=0}^{2} B_j(t)(\mathbf{q}_{2,j} - 2\mathbf{q}_{3,j} + \mathbf{q}_{4,j})\\ =& -4a_1\sum_{j=0}^{2} B_j(t)(\mathbf{q}_{2,j} - \mathbf{q}_{1,j}) + 2\sum_{j=0}^{2} B_j(t)(\mathbf{q}_{0,j} - 2\mathbf{q}_{1,j} + \mathbf{q}_{2,j})\\ =& \frac{\partial^2}{\partial s^2}\mathbf{q}_1(\varphi(0), t).\end{aligned}$$

Therefore $\mathbf{q}_1$ and $\mathbf{q}_2$ join with G^2 continuity. Using an approach similar to that in Theorem 5.1, one easily extends the above proof to cover any depth of subdivision in either parameter. Also at a point where s and t knot lines meet, the approximant is G^2 since each adjacent pair of the four neighbouring patches join with G^2 continuity. This is explained in [12]. The same remark is valid, for example, at the poles of the sphere considered in the next section.

§7. Numerical examples

The approximation scheme was applied to an octant of a sphere with unit radius. $\mathbf{r}$ has weights $w_1 = w_2 = 1/\sqrt{2}$ and control points

$$[\mathbf{p}_{ij}] = \begin{bmatrix} (1,0,0) & (1,0,1) & (0,0,1) \\ (1,1,0) & (1,1,1) & (0,0,1) \\ (0,1,0) & (0,1,1) & (0,0,1) \end{bmatrix}.$$

At each level of subdivision both the error bound E_1 given by (12) and the actual maximum error $E_2 = \sqrt{x^2 + y^2 + z^2} - 1$, found by sampling each patch of the approximant at 20×20 points, are shown in Table 1. Note that E_1 approaches E_2 in the limit, suggesting that the error bound is sharp. Note also that the error at each level is roughly a sixteenth of the previous one.

Figure 4 shows the boundaries of the 16×8 patches of a biquadratic approximant of a whole sphere of radius 1. $\mathbf{r}$ is a rational biquadratic spline with 4×2 patches and there are two levels of subdivision in each direction. The error bound is 3.80×10^{-4}; see also Table 1. Figure 5 shows the boundaries of the 16×16 patches of a biquadratic approximant of a whole torus of outer radius 3 and inner radius 1. $\mathbf{r}$ is a rational biquadratic spline with 4×4 patches and there are again two levels of subdivision in each direction. The error bound is 9.44×10^{-4}.

Number of patches	E_1	E_2
1×1	1.34×10^{-1}	9.70×10^{-2}
2×2	6.49×10^{-3}	5.84×10^{-3}
4×4	3.80×10^{-4}	3.69×10^{-4}
8×8	2.33×10^{-5}	2.31×10^{-5}
16×16	1.45×10^{-6}	1.45×10^{-6}
32×32	9.07×10^{-8}	9.07×10^{-8}

Table 1.

§8. Conclusions

A method for approximating conic sections by quadratic splines with continuous curvature has been presented. Moreover an explicit error bound is derived and this can be used to determine how many subdivisions are required in order to satisfy a given tolerance. The main advantages of this method and the error bound are:

(i) The error bound is $O(h^4)$ which is optimal,

(ii) The spline approximation is both C^1 and G^2, also optimal,

(iii) The scheme applies also to a large number of the most commonly used analytic surfaces in computer-aided design.

There is also a good potential for generalising these ideas to higher degree spline approximations and higher degree rational polynomials [11].

Acknowledgement. I would like to thank Knut Mørken for helpful and enjoyable discussions about high order approximation. I would also like to thank the referees for helpful comments and suggestions for improving the readability of this paper.

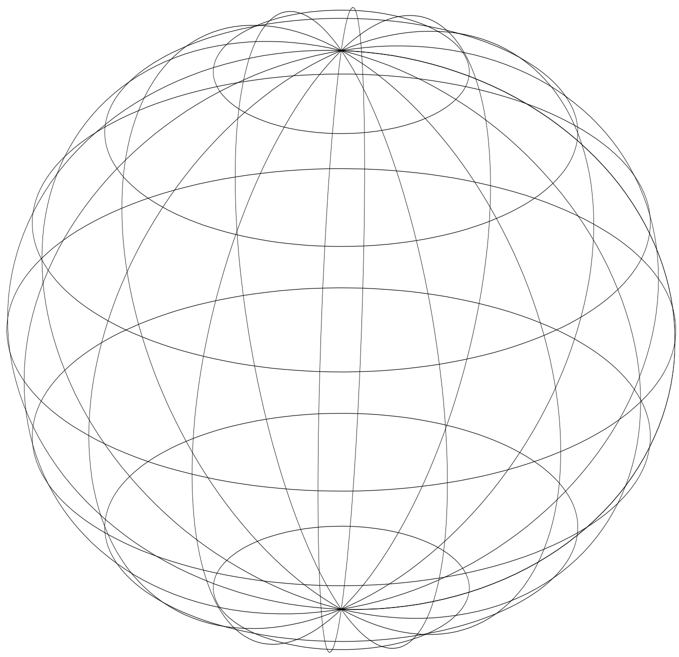

Figure 4. Approximation of a sphere.

§9. References

1. Bardis L., & Patrikalakis M., Approximate conversion of rational B-spline patches, Computer-Aided Geom. Design **6** (1989), 189–204.
2. de Boor, C., *A Practical Guide to Splines*, Springer-Verlag, New York, 1978.
3. de Boor, C., Höllig, K., & Sabin M., High accuracy geometric Hermite interpolation, Computer-Aided Geom. Design **4** (1987), 269–278.
4. Degen, W., High accurate rational approximation of parametric curves, Computer-Aided

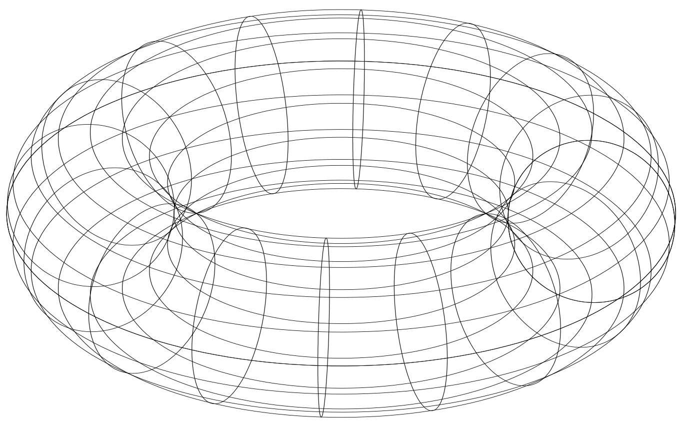

Figure 5. Approximation of a torus.

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. Faux, I., & Pratt, M., *Computational geometry for design and manufacture*, Ellis Horwood, England, 1979.

8. Filip, D., Magedson, R., & Markot, R., Surface algorithms using bounds on derivatives, Computer-Aided Geom. Design **3** (1986), 295–311.
9. Floater, M. S., Derivatives of rational Bézier curves, Computer-Aided Geom. Design **9** (1992), 161–174.
10. Floater, M. S., Evaluation and properties of the derivative of a NURBS curve, in *Mathematical Methods in CAGD, T. Lyche & L. L. Schumaker (eds.), Academic Press, Boston*, (1992), 261–274.
11. Floater, M. S., An $O(h^{2n})$ Hermite approximation for conic sections, *preprint*.
12. Gregory, J. A., Geometric continuity, in *Mathematical Methods in CAGD, T. Lyche & L. L. Schumaker (eds.), Academic Press, Boston*, (1989), 353–371.
13. Hoschek, J., Approximate conversion of spline curves, Computer-Aided Geom. Design **4** (1987), 59–66.
14. Hoschek, J., & Schneider, F., Spline conversion for trimmed rational Bézier- and B-spline surfaces, Comput. Aided Design **22** (1990), 580–590.
15. Lyche, T., Mørken, K., A metric for parametric approximation, preprint.
16. Mørken, K., Best approximation of circle segments by quadratic Bézier curves, in *Curves and Surfaces, P. L. Laurent, A. Le Méhaute & L. L. Schumaker (eds.) Academic Press, Boston*, (1991), 331–336.
17. Patrikalakis, M., Approximate conversion of rational splines, Computer-Aided Geom. Design **6** (1989), 155–165.
18. Piegl, L., & Tiller, W., Curve and surface constructions using rational B-splines, Comput. Aided Design **19** (1987), 485–498.
19. Schaback, R., Interpolation with piecewise quadratic visually C^2 Bézier polynomials, Computer-Aided Geom. Design **6** (1989), 219–233.
20. Sederberg, T. & Kakimoto, M., Approximating rational curves using polynomial curves, in *NURBS for Curve and Surface Design, G. Farin (ed), SIAM, Philadelphia*, (1991), 149–158.