ELSEVIER

Computer Aided Geometric Design 16 (1999) 755–766

COMPUTER AIDED GEOMETRIC DESIGN

G^3 approximation of conic sections by quintic polynomial curves

Lian Fang [1]

Unigraphics Solutions Inc., 10824 Hope Street, Cypress, CA 90630, USA

Received August 1998; revised December 1998

Abstract

This paper presents a method for approximating conic sections using quintic polynomial curves. The constructed quintic polynomial curve has G^3-continuity with the conic section at the end points and G^1-continuity at the parametric mid-point. It is found that for any conic section, there exist three quintic polynomial curves satisfying the mentioned geometric continuity. One of them is the geometric Hermite interpolant proposed in (Floater, 1997) and one of the others is shown to have much smaller error and better shape-preserving property. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Conic sections; Rational Bézier curve; Quintic polynomial; Geometric continuity

1. Introduction

It was described in a recent paper (Floater, 1997), how one can approximate a conic section, in the form of a rational quadratic Bézier curve, by a geometric Hermite interpolant of any odd degree n. The interpolant has a total number of $2n$ contacts with the conic section: $n-1$ at the end points and 2 at the parametric mid-point. Thus, for a quintic polynomial interpolant ($n=5$), it is G^3-continuous with the conic section at the end points and G^1-continuous at the parametric mid-point. Also, in a recent study on circular arc approximation using quintic polynomial curves, it was pointed out that there are three quintic polynomial curves meeting the above-mentioned geometric continuity requirements for circular arcs and Floater's quintic Hermite interpolant is just one of them (Fang, 1998). In this paper we will further prove that in fact this statement is true not only for circular arcs but also for any conic section curves (except parabolas). We will also show that one of these quintic polynomial curves has smaller approximation error and possesses better shape-preserving quality comparing to Floater's quintic Hermite interpolant.

[1] Email: fang@ugsolutions.com.

PII: S0167-8396(99)00017-5

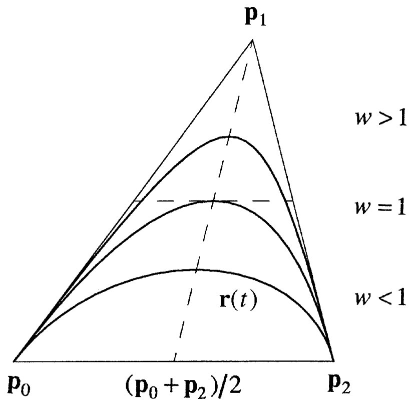

Fig. 1. The conic section $\mathbf{r}(t)$ in the cases $w < 1$, $w = 1$ and $w > 1$.

The problem of interest is stated as follows. Given a conic section represented in the form of a rational quadratic Bézier curve as

$$\mathbf{r}(t) = \frac{B_0(t)\mathbf{p_0} + B_1(t)w\mathbf{p_1} + B_2(t)\mathbf{p_2}}{B_0(t) + B_1(t)w + B_2(t)}, \quad t \in [0, 1], \tag{1}$$

where $\mathbf{p_0}, \mathbf{p_1}, \mathbf{p_2} \in R^2$ are the control points, $w \in R$ is the weight associated with $\mathbf{p_1}$, assumed positive, $B_0(t) = (1-t)^2$, $B_1(t) = 2t(1-t)$, $B_2(t) = t^2$ are the Bernstein basis functions, we want to find an approximating quintic polynomial curve that is G^3-continuous with the conic section at the end points and G^1-continuous at the parametric mid-point. Here the standard form of $\mathbf{r}(t)$ is used without losing any generality. It is also well known that $\mathbf{r}(t)$ is an ellipse when $w < 1$, a parabola when $w = 1$ and a hyperbola when $w > 1$ (see (Farin, 1993)). Fig. 1 shows these three different possibilities.

The rest of this paper is organized as follows. In Section 2 we prove that there exist three quintic polynomial curves satisfying the above-mentioned geometric continuity requirements. Section 3 addresses the approximation quality of the resulted quintic curves. Approximation of circular arcs is given as an example in Section 4 and some concluding remarks are given in Section 5.

2. Conic section approximation

A quintic polynomial curve, when represented in Hermite form, can be written as

$$\mathbf{Q}(t) = (x(t), y(t)) = \begin{bmatrix} H_1(t) & H_2(t) & H_3(t) & H_4(t) & H_5(t) & H_6(t) \end{bmatrix} \begin{bmatrix} \mathbf{Q}(0) \\ \mathbf{Q}'(0) \\ \mathbf{Q}''(0) \\ \mathbf{Q}(1) \\ \mathbf{Q}'(1) \\ \mathbf{Q}''(1) \end{bmatrix},$$

$$t \in [0, 1], \tag{2}$$

where $\mathbf{Q}(0),\ \mathbf{Q}'(0),\ \mathbf{Q}''(0),\ \mathbf{Q}(1),\ \mathbf{Q}'(1)$ and $\mathbf{Q}''(1)$ are the position, the first derivative and the second derivative of $\mathbf{Q}(t)$ at $t=0$ and $t=1$, respectively, and $H_i(t),\ i=1,\dots,6$, are the quintic Hermite polynomial functions, which are listed below without any derivations.

$$\begin{aligned}
H_1(t) &= 1-10t^3+15t^4-6t^5,\\
H_2(t) &= t-6t^3+8t^4-3t^5,\\
H_3(t) &= \tfrac{1}{2}t^2-\tfrac{3}{2}t^3+\tfrac{3}{2}t^4-\tfrac{1}{2}t^5,\\
H_4(t) &= 10t^3-15t^4+6t^5,\\
H_5(t) &= -4t^3+7t^4-3t^5,\\
H_6(t) &= \tfrac{1}{2}t^3-t^4+\tfrac{1}{2}t^5.
\end{aligned}$$

Interested readers are pointed to (Hosaka, 1969) and (Hoschek, 1993) for more details about Hermite polynomials.

It is well known that a quintic Hermite curve described by (2) can be G^2-continuous to $\mathbf{r}(t)$ at end points by setting

$$\begin{aligned}
&\mathbf{Q}(0)=\mathbf{r}(0), \quad \mathbf{Q}'(0)=\alpha_0\mathbf{r}'(0), \quad \mathbf{Q}''(0)=\alpha_0^2\mathbf{r}''(0)+\beta_0\mathbf{r}'(0),\\
&\mathbf{Q}(1)=\mathbf{r}(1), \quad \mathbf{Q}'(1)=\alpha_1\mathbf{r}'(1), \quad \mathbf{Q}''(1)=\alpha_1^2\mathbf{r}''(1)+\beta_1\mathbf{r}'(1),
\end{aligned} \tag{3}$$

where $\alpha_0,\ \beta_0,\ \alpha_1$ and β_1 are arbitrary constants and the differential properties of $\mathbf{r}(t)$ at $t=0$ and $t=1$ are computed as

$$\begin{aligned}
&\mathbf{r}(0)=\mathbf{p}_0, && \mathbf{r}(1)=\mathbf{p}_2,\\
&\mathbf{r}'(0)=2w(\mathbf{p}_1-\mathbf{p}_0), && \mathbf{r}'(1)=2w(\mathbf{p}_2-\mathbf{p}_1),\\
&\mathbf{r}''(0)=\big(4w-8w^2\big)(\mathbf{p}_1-\mathbf{p}_0) && \mathbf{r}''(1)=\big(-4w+8w^2\big)(\mathbf{p}_2-\mathbf{p}_1)\\
&\qquad +2(\mathbf{p}_2-\mathbf{p}_0), && \qquad -2(\mathbf{p}_2-\mathbf{p}_0).
\end{aligned}$$

2.1. The G^3-continuity at $t=0$ and $t=1$

To achieve G^3-continuity at the end points, the third derivatives of $\mathbf{Q}(t)$ at $t=0$ and $t=1$ need to satisfy

$$\mathbf{Q}'''(0)=\alpha_0^3\mathbf{r}'''(0)+3\alpha_0\beta_0\mathbf{r}''(0)+\gamma_0\mathbf{r}'(0), \tag{4a}$$

$$\mathbf{Q}'''(1)=\alpha_1^3\mathbf{r}'''(1)+3\alpha_1\beta_1\mathbf{r}''(1)+\gamma_1\mathbf{r}'(1), \tag{4b}$$

where

$$\mathbf{r}'''(0)=-12(1-w)\big(4w^2(\mathbf{p}_1-\mathbf{p}_0)-(\mathbf{p}_2-\mathbf{p}_0)\big),$$

$$\mathbf{r}'''(1)=-12(1-w)\big(4w^2(\mathbf{p}_2-\mathbf{p}_1)-(\mathbf{p}_2-\mathbf{p}_0)\big),$$

and γ_0 and γ_1 are arbitrary constants.

Substituting $\mathbf{Q}'''(0)$ obtained from (2) and (3) into (4a), taking cross product with $\mathbf{r}'(0)$ on both sides of (4a) to eliminate the γ_0 term, and using the fact that

$$(\mathbf{p}_2-\mathbf{p}_0)\times(\mathbf{p}_1-\mathbf{p}_0),\quad (\mathbf{p}_2-\mathbf{p}_1)\times(\mathbf{p}_1-\mathbf{p}_0)\quad \text{and}\quad (\mathbf{p}_2-\mathbf{p}_1)\times(\mathbf{p}_2-\mathbf{p}_0)$$

all equal two times the area of the triangle $\Delta\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$, (4a) becomes

$$-\alpha_0\beta_0+w\beta_1+\left[10-3\alpha_0^2-2(1-w)\alpha_0^3-8w\alpha_1-\left(1+2w-4w^2\right)\alpha_1^2\right]=0. \tag{5a}$$

Applying similar procedure on (4b) results in

$$w\beta_0-\alpha_1\beta_1-\left[10-3\alpha_1^2-2(1-w)\alpha_1^3-8w\alpha_0-\left(1+2w-4w^2\right)\alpha_0^2\right]=0. \tag{5b}$$

From (5a) and (5b), it is clear that by selecting one set of α_0 and α_1, there exist a unique solution for β_0 and β_1 satisfying (5a) and (5b) provided $\alpha_0\alpha_1-w^2$ is not zero. This means that there are infinite quintic polynomials that are G^3-continuous with $\mathbf{r}(t)$ at end points.

2.2. The G^1-continuity at $t=1/2$

To achieve G^1-continuity at the parametric mid-point, we require $\mathbf{Q}(1/2)=\mathbf{r}(1/2)$ and $\mathbf{Q}'(1/2)\times\mathbf{r}'(1/2)=0$. Because any point $(x,y)\in R^2$ can be written uniquely in terms of barycentric coordinates with respect to a triangle, the position continuity at $t=1/2$ immediately suggests that the barycentric coordinates of $\mathbf{Q}(1/2)$ and $\mathbf{r}(1/2)$ should be the same. Therefore, we have

$$\overline{L}_0=L_0,\quad \overline{L}_1=L_1\quad\text{and}\quad \overline{L}_2=L_2,$$

where $\overline{L}_i$ and L_i $(i=0,1,2)$ are the barycentric coordinates of $\mathbf{Q}(1/2)$ and $\mathbf{r}(1/2)$ with respect to the triangle $\Delta\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$, i.e.,

$$\mathbf{Q}(1/2)=\overline{L}_0\mathbf{p}_0+\overline{L}_1\mathbf{p}_1+\overline{L}_2\mathbf{p}_2$$

and

$$\mathbf{r}(1/2)=L_0\mathbf{p}_0+L_1\mathbf{p}_1+L_2\mathbf{p}_2.$$

Direct computation of $\mathbf{Q}(1/2)$ and $\mathbf{r}(1/2)$ gives

$$\overline{L}_0=\tfrac{1}{2}-\tfrac{5}{16}w\alpha_0-\tfrac{1}{16}\left(w-2w^2\right)\alpha_0-\tfrac{1}{32}w\beta_0-\tfrac{1}{32}\left(\alpha_0^2-\alpha_1^2\right),$$
$$\overline{L}_1=\tfrac{5}{16}w(\alpha_0+\alpha_1)+\tfrac{1}{16}\left(w-2w^2\right)\left(\alpha_0^2+\alpha_1^2\right)+\tfrac{1}{32}w(\beta_0-\beta_1),$$
$$\overline{L}_2=\tfrac{1}{2}-\tfrac{5}{16}w\alpha_1-\tfrac{1}{16}\left(w-2w^2\right)\alpha_1^2+\tfrac{1}{32}w\beta_1+\tfrac{1}{32}\left(\alpha_0^2-\alpha_1^2\right),$$
$$L_0=L_2=\tfrac{1}{2(1+w)}\quad\text{and}\quad L_1=\tfrac{w}{1+w}.$$

Substituting $\overline{L}_i$ and L_i $(i=0,1,2)$ to $\overline{L}_0-\overline{L}_2=L_0-L_2=0$ (since $L_0=L_2$) and $\overline{L}_1=L_1$ yields

$$\tfrac{5}{16}w(\alpha_0-\alpha_1)+\tfrac{1}{16}\left(1+w-2w^2\right)\left(\alpha_0^2-\alpha_1^2\right)+\tfrac{1}{32}w(\beta_0+\beta_1)=0, \tag{6a}$$
$$\tfrac{5}{16}w(\alpha_0+\alpha_1)+\tfrac{1}{16}\left(w-2w^2\right)\left(\alpha_0^2+\alpha_1^2\right)+\tfrac{1}{32}w(\beta_0-\beta_1)=\tfrac{w}{1+w}. \tag{6b}$$

Furthermore, the tangency continuity at $t=1/2$ results in

$$\tfrac{7}{16}(\alpha_0-\alpha_1)+\tfrac{1}{16}(1-2w)\left(\alpha_0^2-\alpha_1^2\right)+\tfrac{1}{32}(\beta_0+\beta_1)=0. \tag{7}$$

Subtracting (7) $\times\, w$ from (6a) gives

$$\tfrac{1}{16}(\alpha_0-\alpha_1)(\alpha_0+\alpha_1-2w)=0 \tag{8}$$

which shows that for the quintic polynomial $\mathbf{Q}(t)$ to be G^1-continuous with $\mathbf{r}(t)$ at $t=1/2$, α_0 and α_1 need to satisfy either $\alpha_0-\alpha_1=0$, $\alpha_0+\alpha_1-2w=0$, or both.

So far, we have shown that (5a) and (5b) need to be met for $\mathbf{Q}(t)$ to be G^3-continuous with $\mathbf{r}(t)$ at end points and that (6b) and (8) need to be met for $\mathbf{Q}(t)$ to be G^1-continuous with $\mathbf{r}(t)$ at $t=1/2$. In the following we will show that there exist only three solutions satisfying all of the four equations.

Assuming $\alpha_0-\alpha_1=0$, thus $\beta_0+\beta_1=0$, we let $\alpha_0=\alpha_1=\alpha$ and $\beta_0=-\beta_1=\beta$ to simplify notation, then (4a) and (4b) become identical as

$$(-2+2w)\alpha^3+\bigl(4w^2-2w-4\bigr)\alpha^2-(8w+\beta)\alpha-w\beta+10=0, \tag{9}$$

and (6b) becomes

$$\frac{5}{8}w\alpha+\frac{1}{8}\bigl(w-2w^2\bigr)\alpha^2+\frac{1}{16}w\beta=\frac{w}{1+w},\quad \text{or}$$
$$\beta=\frac{16}{1+w}-10\alpha-2(1-2w)\alpha^2. \tag{10}$$

Substituting (10) into (9) yields

$$\begin{aligned}&-2w\alpha^3+6\alpha^2+\frac{2(w^2+w-8)}{1+w}\alpha-\frac{2(3w-5)}{1+w}\\ &\quad=-2(\alpha-1)\left[w\alpha^2+(w-3)\alpha+\frac{5-3w}{1+w}\right]=0,\end{aligned} \tag{11}$$

for which the solutions for α can be easily found as

$$1,\quad \overline{\alpha}_1=\frac{1}{2w}\left((3-w)-(1-w)\sqrt{\frac{w+9}{w+1}}\right)\quad\text{and}$$
$$\overline{\alpha}_2=\frac{1}{2w}\left((3-w)+(1-w)\sqrt{\frac{w+9}{w+1}}\right). \tag{12}$$

Distributions of these three solutions as functions of w are shown in Fig. 2. It should be noted that when $w=1$, (11) has triple roots at $\alpha=1$.

If $\alpha_0+\alpha_1-2w=0$, thus $\beta_0+\beta_1=(8w^2-4w-14)(\alpha_0-\alpha_1)$, we have

$$\alpha_1=2w-\alpha_0 \tag{13a}$$

and

$$\beta_1=2(8w^2-4w-14)(\alpha_0-w)-\beta_0. \tag{13b}$$

Substituting them into (6b) yields

$$\beta_0=(4w-2)\alpha_0^2-14\alpha_0+\frac{4(w^2+w+4)}{1+w}, \tag{14}$$

and (4a) and (4b) become

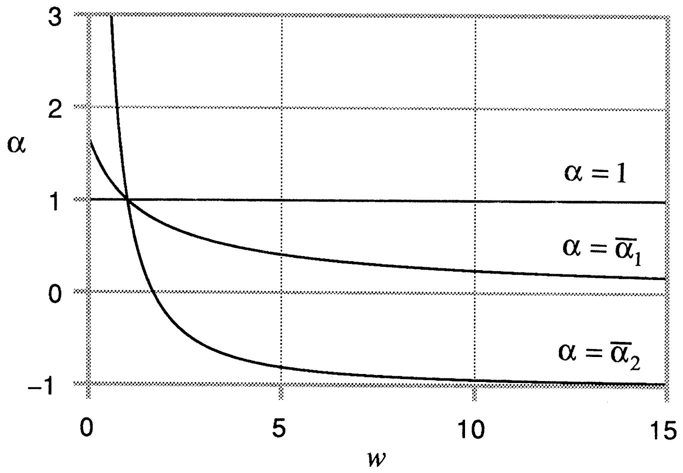

Fig. 2. The graphs of α as functions of w.

$$-2w\alpha_0^3 + 10\alpha_0^2 - \frac{2}{1+w}\big(3w^2 + 3w + 8\big)\alpha_0 + \frac{2}{1+w}\big(2w^3 + 2w^2 - 3w + 5\big) = 0, \tag{15a}$$

$$-2w\alpha_0^3 + \big(12w^2 - 10\big)\alpha_0^2 - \frac{2}{1+w}\big(12w^4 + 12w^3 - 17w^2 - 17w + 8\big)\alpha_0 + \frac{2}{1+w}\big(8w^5 + 8w^4 - 16w^3 - 16w^2 + 19w - 5\big) = 0. \tag{15b}$$

These two equations are found to have common solution $\alpha_0 = 1$ only when $w = 1$. However, this solution is included in the solutions found previously.

With the above derivation, we can conclude that for any given conic section represented by (1), as long as $w \neq 1$, there exist three quintic polynomials that are G^3-continuous with the conic section at the end points and G^1-continuous at the parametric mid-point. Therefore, from (2) and letting $\alpha_0 = \alpha_1 = \alpha$ and $\beta_0 = -\beta_1 = \beta$, the approximating quintic curve $\mathbf{Q}(t)$ can be rewritten as

$$\mathbf{Q}(t) = \overline{K}_0(t)\mathbf{p}_0 + \overline{K}_1(t)\mathbf{p}_1 + \overline{K}_2(t)\mathbf{p}_2, \tag{16}$$

where

$$\overline{K}_0(t) = H_1(t) - 2w\alpha H_2(t) + \big[-\big(2 + 4w - 8w^2\big)\alpha^2 - 2w\beta\big]H_3(t) + 2\alpha^2 H_6(t),$$
$$\overline{K}_1(t) = 2w\alpha(H_2(t) - H_5(t)) + \big[\big(4w - 8w^2\big)\alpha^2 + 2w\beta\big](H_3(t) + H_6(t)),$$
$$\overline{K}_2(t) = H_4(t) + 2w\alpha H_5(t) + \big[-\big(2 + 4w - 8w^2\big)\alpha^2 - 2w\beta\big]H_6(t) + 2\alpha^2 H_3(t),$$

where α equals 1, $\overline{\alpha}_1$, or $\overline{\alpha}_2$ and β is obtained from (10). Note that $\overline{K}_0(t) + \overline{K}_1(t) + \overline{K}_2(t) = 1$ reflects that they are indeed the barycentric coordinates of $\mathbf{Q}(t)$ with respect to the triangle $\Delta\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$. For convenience, we denote the three quintic approximation curves obtained by using $\alpha = 1,\ \overline{\alpha}_1$, and $\overline{\alpha}_2$ as $\mathbf{Q}_0(t),\ \mathbf{Q}_1(t)$ and $\mathbf{Q}_2(t)$, respectively. In the next section, we will evaluate the quality of these three quintic approximation curves.

3. Approximation accuracy

3.1. Approximation error

It is well known that for any point on the conic section $\mathbf{r}(t)$, its barycentric coordinates $\tau_0,\ \tau_1,\ \tau_2$, where $\tau_0 + \tau_1 + \tau_2 = 1$, with respect to the triangle $\Delta\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$, satisfies

$$f\bigl(\mathbf{r}(t)\bigr) = \tau_1^2(t) - 4w^2\tau_0(t)\tau_2(t) = 0. \tag{17}$$

Consequently, for any curve $\mathbf{c}(t)$ approximating the conic section $\mathbf{r}(t)$, we can use $f(\mathbf{c}(t))$ to see how well the approximation is. The function $f(\mathbf{Q}(t)) = \overline{K}_1^2(t) - 4w^2\overline{K}_0(t)\overline{K}_2(t)$ is computed as

$$f\bigl(\mathbf{Q}_i(t)\bigr) = A_i t^4(1-t)^4(2t-1)^2, \quad i = 0, 1, 2, \tag{18}$$

where

$$A_0 = \frac{16w^2(w-1)^4}{(w+1)^2}, \quad A_1 = \frac{2(w-1)^4\delta_-(w)}{w^2(w+1)}, \quad A_2 = \frac{2(w-1)^4\delta_+(w)}{w^2(w+1)} \quad \text{and}$$

$$\delta_\mp(w) = (w+1)(2w+9)(2w+3)\left(1 \mp \sqrt{\frac{w+9}{w+1}}\right) + 2(w+3)(8w+9).$$

It can be shown that $0 \leqslant A_1 \leqslant A_0 \leqslant A_2$ and the equality happens when $w = 1$.

Combining (12), (16) and (18), we can deduce the following remarks.

Remark 1. All the three approximation curves are always "outside" the conic section, i.e., $\mathbf{Q}_i(t)$ $(i = 0, 1, 2)$ always lies on the side containing $\mathbf{p}_1$. This is a direct result from the fact that $f(\mathbf{Q}_i(t))$ is always nonnegative.

Remark 2. $\mathbf{Q}_0(t)$, the approximation curve obtained by using $\alpha = 1$, is the same as the quintic Hermite interpolant proposed in (Floater, 1997). This can be easily verified by comparing $\mathbf{Q}_0(t)$ against the result in (Floater, 1997) with $n = 5$.

Remark 3. $\mathbf{Q}_2(t)$, the approximation curve obtained by using $\alpha = \overline{\alpha}_2$, is generally unacceptable because $\overline{\alpha}_2$ becomes negative when w is greater than $5/3$ and approaches infinity when w approaches zero.

Remark 4. As shown in Fig. 3, the ratio $\lambda = f(\mathbf{Q}_1(t))/f(\mathbf{Q}_0(t)) = A_1/A_0$ decreases exponentially as w increases. The Taylor expansion of λ at $w = 0$ also shows that λ approaches $1/81$ when w approaches zero.

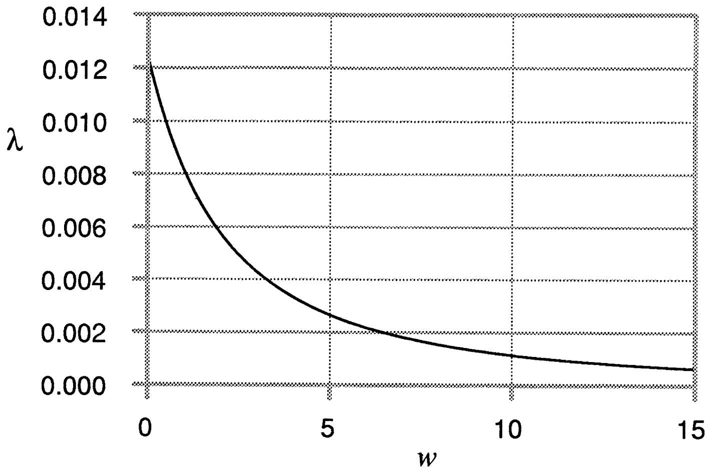

Fig. 3. The graph of the error ratio λ as a function of w.

Remark 5. $\mathbf{Q}_1(t)$, the approximation curve obtained by using $\alpha = \overline{\alpha}_1$, is completely enclosed by $\mathbf{Q}_0(t)$ and $\mathbf{r}(t)$ and these three curves intersect only at $t = 0,\ 1/2$ and 1.

Remark 6. For $0 < w \leqslant 3$,

$$d_H(\mathbf{Q}_1, \mathbf{r}) \leqslant \frac{4}{3125} \frac{\max(1, w^2)(w-1)^4\lambda}{(1+w)^2} \left|\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2\right|, \tag{19}$$

where $d_H(\mathbf{Q}_1, \mathbf{r})$ is the Hausdorff distance between $\mathbf{Q}_1(t)$ and $\mathbf{r}(t)$ and $\lambda = A_1/A_0$.

To prove this theorem, we will use several lemmas proven in (Floater, 1997). First, from Lemmas 3.3a and 3.3b of (Floater, 1997), we know that $\mathbf{Q}_0(t)$ lies entirely inside the triangle $\Delta\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$ for $0 < w \leqslant 3$. Since $\mathbf{Q}_1(t)$ is always enclosed by $\mathbf{Q}_0(t)$ and $\mathbf{r}(t)$ (from Remark 5), $\mathbf{Q}_1(t)$ will also lie entirely inside the triangle $\Delta\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$ for $0 < w \leqslant 3$. Therefore, from Lemma 3.2 of (Floater, 1997), we obtain

$$d_H(\mathbf{Q}_1, \mathbf{r}) \leqslant \frac{1}{4} \max\left(\frac{1}{w^2}, 1\right) \max_{t\in[0,1]} \left|f(\mathbf{Q}_1(t))\right| \left|\mathbf{p}_0 - 2\mathbf{p}_1 + \mathbf{p}_2\right| \quad \text{for } 0 < w \leqslant 3. \tag{20}$$

Furthermore, from Lemma 3.4 of (Floater, 1997), we have

$$f\bigl(\mathbf{Q}_0(t)\bigr) \leqslant \frac{16}{3125} \frac{w^2(w-1)^4}{(1+w)^2} \quad \text{for all } t \in [0, 1]. \tag{21}$$

Combining (20), (21) and $f(\mathbf{Q}_1(t)) = \lambda f(\mathbf{Q}_0(t))$, we obtain (19) as claimed.

It should be noted that (19) is valid not only when $0 < w \leqslant 3$, it remains true as long as $\mathbf{Q}_1(t)$ is entirely inside the triangle $\Delta\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$. Given the facts that $\mathbf{Q}_0(t)$ starts to migrate outside the triangle when $w > 3$ and $f(\mathbf{Q}_1(t))$ is much smaller than $f(\mathbf{Q}_0(t))$, it is reasonable to assume that $\mathbf{Q}_1(t)$ will remain inside the triangle $\Delta\mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$ even w is greater than 3 (but smaller than a certain value). Therefore, (19) should be valid for $0 < w \leqslant w_0$ where $w_0 > 3$.

3.2. Shape preservation

The approximation quality is also evaluated by checking whether $\mathbf{Q}(t)$ preserve the shape of $\mathbf{r}(t)$. An important property of $\mathbf{r}(t)$ is that $\mathbf{r}(1/2)$ is the only point at which the tangent of $\mathbf{r}(t)$ is parallel to line $\overline{\mathbf{p}_0\mathbf{p}_2}$. Similarly, when $\mathbf{Q}(1/2)$ is the only point at which the tangent of $\mathbf{Q}(t)$ is parallel to line $\overline{\mathbf{p}_0\mathbf{p}_2}$, we say that $\mathbf{Q}(t)$ preserves the basic shape of $\mathbf{r}(t)$. This criterion can be written mathematically as follows.

Criterion 1. The equation

$$\frac{\mathrm{d}\mathbf{Q}(t)}{\mathrm{d}t} \times (\mathbf{p}_2 - \mathbf{p}_0) = 0 \tag{22}$$

has only one solution at $t = 1/2$ for $t \in [0, 1]$.

Direct computation of (22) gives

$$\begin{aligned} &2w\left[\left(16\alpha - \frac{32}{1+w}\right)t^3 - \left(24\alpha - \frac{48}{1+w}\right)t^2 \right.\\ &\quad \left. + \left(10\alpha - \frac{16}{1+w}\right)t - \alpha\right](\mathbf{p}_2 - \mathbf{p}_1) \times (\mathbf{p}_2 - \mathbf{p}_0) \\ &= 16w(2t-1)\left[\left(\alpha - \frac{2}{1+w}\right)t^2 \right.\\ &\quad \left. - \left(\alpha - \frac{2}{1+w}\right)t + \frac{\alpha}{8}\right](\mathbf{p}_2 - \mathbf{p}_1) \times (\mathbf{p}_2 - \mathbf{p}_0) = 0. \end{aligned} \tag{23}$$

Since w and $(\mathbf{p}_2 - \mathbf{p}_1) \times (\mathbf{p}_2 - \mathbf{p}_0)$ are not zero, satisfaction of Criterion 1 will require the quadratic equation

$$\left(\alpha - \frac{2}{1+w}\right)t^2 - \left(\alpha - \frac{2}{1+w}\right)t + \frac{\alpha}{8} \tag{24}$$

to have either no real roots (two complex roots), two real roots both of which are outside $[0, 1]$, or double roots at $t = 1/2$.

As a result, we find that as long as $0 < (w+1)\alpha \leqslant 4$, (22) has only one solution at $t = 1/2$. Therefore, $\mathbf{Q}_0(t)$, obtained by using $\alpha = 1$, will preserve the basic shape of $\mathbf{r}(t)$ only when $0 < w \leqslant 3$. When $\alpha = 1$ and $w > 3$, (22) will have three real roots that are inside $[0, 1]$, which implies that there exist three locations on $\mathbf{Q}_0(t)$ where the tangent is parallel to $\overline{\mathbf{p}_0\mathbf{p}_2}$. In such cases, $\mathbf{Q}_0(t)$ will exhibit anomalies such as "camel humps" and even loops as shown in Fig. 4.

When $\alpha = \overline{\alpha}_1$, it can be easily shown that $(w+1)\overline{\alpha}_1$ is always smaller than 4, which means that $\mathbf{Q}_1(t)$ will always preserve the basic shape of $\mathbf{r}(t)$. This good shape-preserving property can be seen in Fig. 5 which shows a close-up of $\mathbf{r}(t)$ (with $w = 100$) and $\mathbf{Q}_1(t)$ around $t = 1/2$. In fact, when w approaches infinity and $\mathbf{r}(t)$ becomes two straight lines defined by $\mathbf{p}_0$, $\mathbf{p}_1$ and $\mathbf{p}_2$, $\overline{\alpha}_1$ is the only solution that will result in a converged approximation curve, whose control points $\mathbf{B}_i$ $(i = 0, \ldots, 5)$ are given by

$$\begin{aligned} &\mathbf{B}_0 = \mathbf{p}_0, \quad \mathbf{B}_1 = \mathbf{p}_1 + \tfrac{1}{5}(\mathbf{p}_1 - \mathbf{p}_0), \quad \mathbf{B}_2 = \mathbf{B}_3 = \mathbf{p}_1, \\ &\mathbf{B}_4 = \mathbf{p}_1 - \tfrac{1}{5}(\mathbf{p}_2 - \mathbf{p}_1), \quad \mathbf{B}_5 = \mathbf{p}_2. \end{aligned}$$

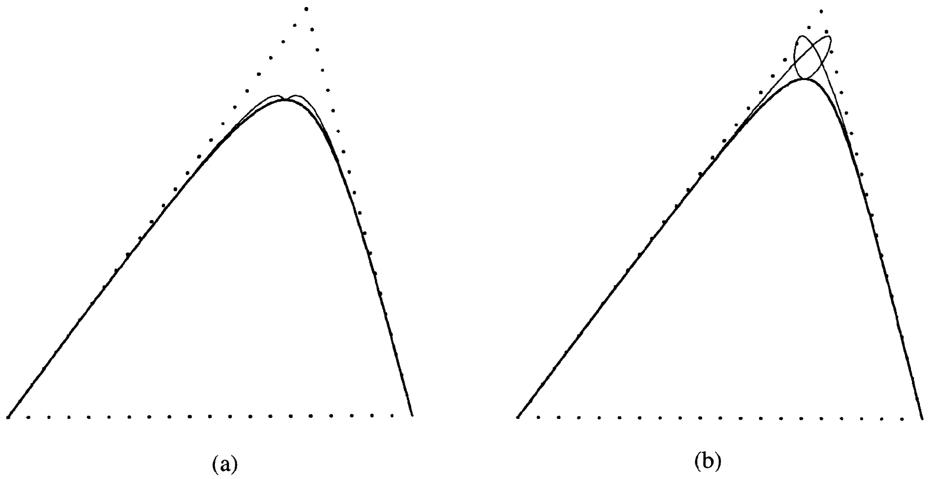

Fig. 4. The anomalies of $\mathbf{Q}_0(t)$ (in thin lines) when $w > 3$: (a) "camel humps" and (b) loops.

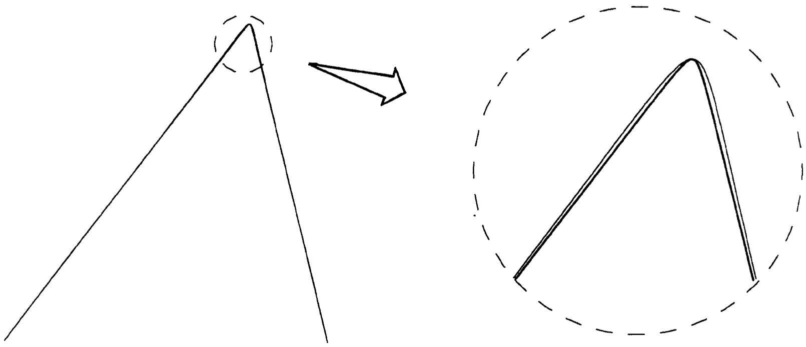

Fig. 5. A close-up of $\mathbf{r}(t)$ ($w = 100$, in thick line) and $\mathbf{Q}_1(t)$ (in thin line) around $t = 1/2$.

Besides preserving the basic shape of $\mathbf{r}(t)$, $\mathbf{Q}_1(t)$ also enjoys better quality in terms of the curvature distribution. For example, Fig. 6 shows that while $\mathbf{Q}_1(t)$ successfully preserve the curvature characteristics of $\mathbf{r}(t)$ (with $w = 2.7$) which has a single peak in its curvature distribution, $\mathbf{Q}_0(t)$ fails because it has two peaks and one valley in its curvature distribution. The numeric results also show that $\mathbf{Q}_1(t)$ will maintain the convexity (free of inflection points) until w is approximately 23 while $\mathbf{Q}_0(t)$ is free of inflection points only when $w < 3$.

4. Example: circular arcs

When $\|\overline{\mathbf{p}_0\mathbf{p}_1}\| = \|\overline{\mathbf{p}_1\mathbf{p}_2}\|$ and $w = \cos\theta$ where θ is the angle $\angle \mathbf{p}_1\mathbf{p}_0\mathbf{p}_2$, the conic section described by (1) becomes a circular arc of angular span 2θ. For a unit circular

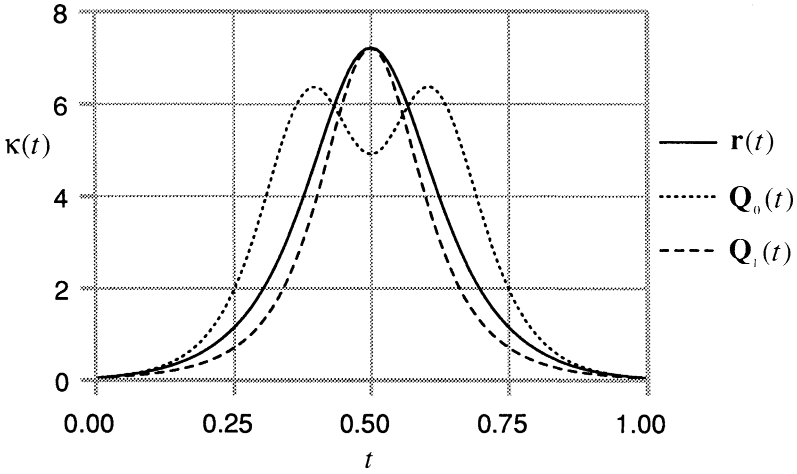

Fig. 6. A sample curvature graph for $\mathbf{r}(t)$ $(w = 2.7)$ and its approximation curves $\mathbf{Q}_0(t)$ and $\mathbf{Q}_1(t)$.

arc with arc's center located at the origin, $f(\mathbf{Q}(t))$ can be related to the radial error $\phi(t) = x^2(t) + y^2(t) - 1$ by

$$f\big(\mathbf{Q}(t)\big) = \frac{\cos^2\theta}{\sin^2\theta}\big(x^2(t) + y^2(t) - 1\big) = \frac{\cos^2\theta}{\sin^2\theta}\phi(t). \tag{25}$$

Thus, the two quintic curves, obtained by using $\alpha = 1$ and $\alpha = \overline{\alpha}_1$, approximating the unit circular arc would have radial errors as

$$\phi_0(t) = \frac{16(1-\cos\theta)^5}{1+\cos\theta}t^4(1-t)^4(2t-1)^2 \quad \text{when } \alpha = 1 \text{ and} \tag{26a}$$

$$\phi_1(t) = \frac{2(1-\cos\theta)^5\delta_-}{\cos^4\theta}t^4(1-t)^4(2t-1)^2 \quad \text{when } \alpha = \overline{\alpha}_1. \tag{26b}$$

In (Fang, 1998), these two quintic curves, represented in Bézier form, and their radial errors were also derived. However, $\phi_1(t)$ was not expressed as an explicit equation. This is because the circular arc was represented by trigonometric functions instead of rational polynomials, which makes it difficult to simplify $\phi_1(t)$ into a concise form.

5. Conclusion

Unlike in (Floater, 1997) in which the geometric Hermite interpolant is given for any odd degree n, we only focus on quintic polynomials in this paper. We proved that for any conic section (except parabolas) described by (1) there exist three quintic polynomial curves that are G^3-continuous with the conic section at end points and G^1-continuous at the parametric mid-point. Among these three quintic curves, one is the same as the quintic interpolant presented in (Floater, 1997) and one of the other two new curves is proven to have much smaller approximation error and better shape-preserving property.

This new quintic approximation curve can be used to construct a spline approximation of $\mathbf{r}(t)$ with C^1 and G^4-continuity by recursively subdividing $\mathbf{r}(t)$ using the scheme discussed in (Floater, 1995), and creating a new approximation curve for each subcurve. It can also be used in approximating rational tensor-product biquadratic Bézier surfaces such as spheres and torus.

Based on the ideas presented above, it may be possible to find polynomials of odd degree n, other than Floater's Hermite interpolants, that are G^{n-2}-continuous with the conic section at end points and G^1-continuous at the parametric mid-point. (In fact, we can easily verify that there exist one unique solution when $n = 3$.) However, the computation involved for higher degree polynomials ($n \geqslant 7$) will be much more complicated and the result might need to be derived numerically.

Acknowledgement

The author would like to thank my colleague Dr. Hang Du who kindly lets me use his algebraic manipulation software: MAPLE for some of the computation presented above. The reviewers' comment and suggestions are also greatly appreciated.

References

Fang, L. (1998), Circular arc approximation by quintic polynomial curves, Computer Aided Geometric Design 15, 843–861.

Farin, G. (1993), *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide,* 3rd ed., Academic Press, San Diego.

Floater, M.S. (1995), High order approximation of conic sections by quadratic splines, Computer Aided Geometric Design 12, 617–637.

Floater, M.S. (1997), An $O(h^{2n})$ Hermite approximation for conic sections, Computer Aided Geometric Design 14, 135–151.

Hosaka, M. (1969), Theory of curve and surface synthesis and their smooth fitting, Information Processing in Japan 9, 60–68.

Hoschek, J. and Lasser, D. (1993), *Fundamentals of Computer Aided Geometric Design,* A.K. Peters.