Regular algebraic curve segments (II)—Interpolation and approximation

Guoliang Xu [a,1], Chandrajit L. Bajaj [b,*], Chuan I Chu [c,2]

[a] *State Key Laboratory of Scientific and Engineering Computing, ICMSEC, Chinese Academy of Sciences, Beijing, People's Republic of China*
[b] *Department of Computer Science, University of Texas, Austin, TX 78712, USA*
[c] *Department of Mathematics, Hong Kong Baptist University, Kowloon, Hong Kong*

Abstract

In this paper (part two of the trilogy) we introduce three classes of reduced form *D*-regular algebraic curve splines and use them for interpolation and approximation of various algebraic curves. Explicit formulas for interpolation and approximation are also given in some low degree cases. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Algebraic curve; Discriminating family; Transversal family; BB-form; Reduced form; Interpolation and approximation

1. Introduction

In the first part of this trilogy of papers (Xu et al., 2000), we introduced a concept of the discriminating family of curves by which regular algebraic curve segments are isolated. Using different discriminating families, several characterizations of the Bernstein–Bézier (BB) form of the implicitly defined real bivariate polynomials over the plane triangle and the quadrilateral are given, so that the zero contours of the polynomials define smooth and connected real algebraic (called regular) curve segments. In the part II of the trilogy of papers, *we use the reduced form regular algebraic curve segments to interpolate or approximate a given curve*. By reduced form, we mean that most of the coefficients of the

BB form of the curve are zeros, which makes the interpolation and approximation problem by the zero contour of the bivariate polynomial as easy as the problem of the univariate polynomials or the rational functions.

It is well known that data interpolation and function approximation by univariate polynomials or rational functions are classical and very well developed areas in the fields of approximation theory as well as CAGD. In contrast, the interpolation and approximation by the zero sets of bivariate polynomials is rather new (Bajaj and Xu, 1996, 1999, 2000; Paluszny and Patterson, 1992, 1993, 1998; Sederberg et al., 1985) and relatively fewer results are available at the present time. After overcoming the difficulties of having singularities and discontinuity properties of real implicitly defined curves in part I of this trilogy of papers, our main results in part II are the existence and uniqueness of the interpolation and approximation solutions, and explicit formulas for solutions. The reduced form real algebraic curves used in this paper have the following advantages:

(a) They are non-singular and without discontinuities.
(b) The curve (that is the polynomial coefficients) can be generated as easily as the classical univariate polynomial or rational function.
(c) However, unlike the polynomial rational functions, the curves are always located in the given domain. Hence the approximation errors are controllable within the domain of interest. This is significant in contrast with the Runge phenomenon of the classical polynomial interpolation (see Example 5.1).
(d) Explicit parameterized expressions exist for the curve that could be used to display the curve efficiently. Hence we have both the implicit form and parametric form. This is very important since we can take the advantage of dual forms.

The solution of our interpolation and approximation problem involves the following steps (see Fig. 1):

Step 1. Build a transform M that converts the problem on triangle or quadrilateral in the xy-plane into a classical rational interpolation or approximation problem on the strip $[0, 1] \times (-\infty, \infty)$ in the st-plane.

Step 2. Solve the classical problem in the st-plane.

Step 3. Transform the solution in the st-plane back to the xy-plane by the inverse transform M^{-1}, to get a solution of the original problem.

Since there are plenty of methods that could be used for solving the classical rational interpolation or approximation problem, the main tasks in these steps are to build the transform M, and to define the proper approximating and approximated curve classes so that the interpolation and approximation problem can be transformed into the classical rational one.

The rest of the paper is organized as follows: Transform M is constructed in Section 2, using a discriminating family, which is introduced in paper I, and a transversal family, which is introduced in this paper. In Section 3, we introduce the approximating curve classes, which are the reduced form D-regular curve families. Then in Section 4, we determine the corresponding approximated curve classes. In Section 5, we convert the
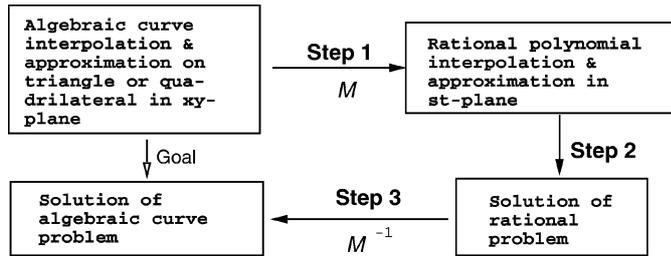
Fig. 1. The pipeline of solving the problem of interpolation and approximation by D-regular algebraic curves.

interpolation and approximation problems to the classical ones with transform M, and hence every problem proposed has a rational polynomial-like solution. Parametric form formulas for the solutions of the interpolation and approximation problems are also provided in this section. The last section concludes the paper and an example that shows partially the significance of the paper is given.

The notations used in this part will follow the first part of the papers. That is, for a general description, we use the xy coordinate system. For a concrete case, such as the discussion on a triangle, we shall use the local barycentric coordinate system and the BB-form for polynomials. On a quadrilateral, we shall use the local uv coordinate system and the tensor product Bernstein form polynomials.

2. Transformation

To establish a map from a triangle or quadrilateral in the xy-plane to the strip $[0, 1] \times (-\infty, \infty)$ in the st-plane, we introduce the concept of a transversal family corresponding to the concept of a discriminating family. A discriminating family and a transversal family together form a pair which serves as a coordinate system within the domain of interest. First we repeat the definition of discriminating family from (Xu et al., 2000).

Definition 2.1. For a given triangle or quadrilateral R, let R_1 and R_2 be two closed pieces of boundary of R with $R_1 \cap R_2 = \emptyset$ (see Fig. 2). Let $D = \{A_s(x, y) = \gamma(x, y) - s\delta(x, y) = 0 \colon s \in [0, 1]\}$ be an algebraic curve family with s as a parameter and $\delta(x, y) > 0$ on $R \setminus \{R_1, R_2\}$ such that
  (1) Each curve in D passes through R_1 and R_2.
  (2) Each curve in D is regular in the interior of R.
  (3) For $\forall p \in R \setminus \{R_1, R_2\}$, there exists one and only one $s \in [0, 1]$ such that $A_s(p) = 0$.
Then we say D is a discriminating family on R, denoted by $D(R, R_1, R_2)$.

Definition 2.2. Let $D(R, R_1, R_2)$ be a given discriminating family, and $T(R, R_1', R_2') = \{B_t(x, y) = \mu(x, y) - t\nu(x, y) = 0 \colon t \in (-\infty, \infty)\}$ be an algebraic curve family with t being a linear parameter, $\mu(x, y)$ and $\nu(x, y)$ are bivariate polynomials and $\nu(x, y) > 0$ on $R \setminus \{R_1, R_2\}$ and R_1' and R_2' being two open (no end points) pieces of the boundary ∂R of R (see Fig. 2). If
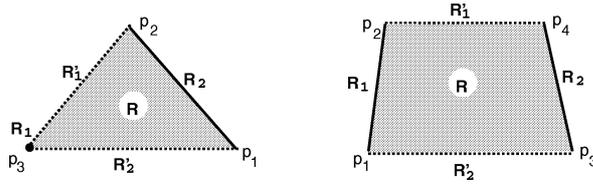
Fig. 2. R_1, R_2, R_1', R_2' for a triangle and a quadrilateral.
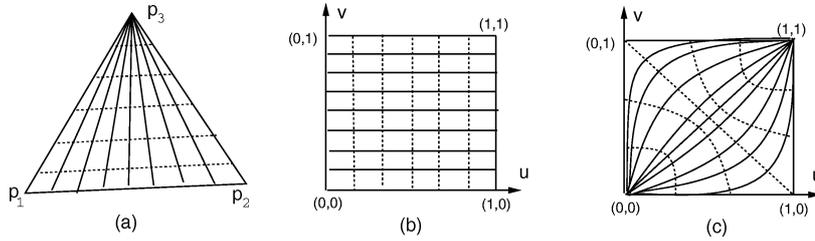
Fig. 3. Discriminating families (real lines) and transversal families (dotted lines).

(1) $\partial R \setminus (R_1 \cup R_2) = R_1' \cup R_2'$ and $R_1' \cap R_2' = \emptyset$;
(2) Each curve in T passes through R_1' and R_2';
(3) Each curve in T is $D(R, R_1, R_2)$-regular;
(4) For $\forall p \in R \setminus \{R_1, R_2\}$, there exists one and only one $t \in (-\infty, \infty)$ such that $B_t(p) = 0$.

Then we say $T(R, R_1', R_2')$ is a transversal family of $D(R, R_1, R_2)$.

Example 2.1. Let

$$T_1\big([p_1 p_2 p_3], (p_1 p_3), (p_2 p_3)\big)$$
$$= \big\{ B_1^2(\alpha_3)t - B_2^2(\alpha_3) + B_0^2(\alpha_3) = 0 \colon t \in (-\infty, \infty) \big\}.$$

Then $T_1([p_1 p_2 p_3], (p_1 p_3), (p_2 p_3))$ is a transversal family of $D_1([p_1 p_2 p_3], p_3, [p_1 p_2])$ (see Fig. 3(a)), where $\alpha = (\alpha_1, \alpha_2, \alpha_3)^{\mathrm{T}}$ are the barycentric coordinates with respect to the triangle $[p_1 p_2 p_3]$. This notation is used throughout the paper.

Example 2.2. Let

$$T_2\big([p_1 p_2 p_3], (p_1 p_2), (p_1 p_3)\big) = \big\{ (\alpha_1^2 + \alpha_2 \alpha_3)t - \alpha_3^2 + \alpha_2^2 = 0 \colon t \in (-\infty, \infty) \big\}.$$

Then $T_2([p_1 p_2 p_3], (p_1 p_2), (p_1 p_3))$ is a transversal family of $D_2([p_1 p_2 p_3], p_2, p_3)$.

Example 2.3. Let

$$T_3\big([p_1 p_2 p_3 p_4], (p_1 p_3), (p_2 p_4)\big) = \big\{ B_1^2(u)t - B_2^2(u) + B_0^2(u) = 0 \colon t \in (-\infty, \infty) \big\}.$$

Then $T_3([p_1 p_2 p_3 p_4], (p_1 p_3), (p_2 p_4))$ is a transversal family of $D_3([p_1 p_2 p_3 p_4], [p_1 p_2], [p_3 p_4])$ (see Fig. 3(b)).
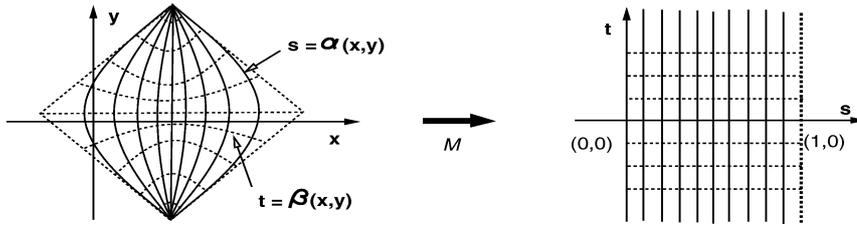
Fig. 4. A discriminating family and its transversal family define (s, t)-coordinate system.

Example 2.4. Let

$$T_4\big([p_1 p_2 p_3 p_4], (p_1 p_2] \cup [p_2 p_4), (p_1 p_3] \cup [p_3 p_4]\big)$$
$$= \big\{[u(1-v) + (1-u)v]t - (1-u-v) = 0 : t \in (-\infty, \infty)\big\}.$$

Then $T_4([p_1 p_2 p_3 p_4], (p_1 p_2] \cup [p_2 p_4), (p_1 p_3] \cup [p_3 p_4])$ is a transversal family of $D_4([p_1 p_2\ p_3 p_4], p_1, p_4)$ (see Fig. 3(c)).

Note that T_1 and T_3 consist of straight lines. One may ask why they should not be defined linearly by $t = \alpha_3$ or $t = u$. The reason for using a quadratic is that we require $t \in (-\infty, \infty)$ for α_3 or u in $[0, 1]$.

Given a discriminating family $D = \{A_s(x, y) = \gamma(x, y) - s\delta(x, y) = 0\}$ and its transversal family $T = \{B_t(x, y) = \mu(x, y) - tv(x, y) = 0\}$, we are in fact given a map between $R \setminus \{R_1, R_2\}$ in xy-plane and the strip $[0, 1] \times (-\infty, \infty)$ in the st-plane (see Fig. 4). Since s and t are linear parameters in $A_s(x, y) = 0$ and $B_t(x, y) = 0$, respectively, they can be written as

$$M(D, T): \begin{cases} s = \alpha(x, y) := \gamma(x, y)/\delta(x, y), \\ t = \beta(x, y) := \mu(x, y)/v(x, y), \end{cases} \tag{2.1}$$

where α and β are well defined rational functions on $R \setminus \{R_1, R_2\}$. We shall denote $M(D_i, T_i)$ as M_i to simplify the notation. Using these transforms, our interpolation or approximation problems in xy-plane will be transformed into the st-plane. In the following examples, we shall give the explicit forms of the transforms for the pairs of families given in the above examples. Furthermore, the inverse transforms are also provided. These inverse transforms are as important as the original ones because the results derived in the st-plane will finally be converted into the xy-plane by them (see Fig. 1).

Example 2.5.

$$M_1: \begin{cases} s = \alpha_2/(\alpha_1 + \alpha_2), \\ t = [B_2^2(\alpha_3) - B_0^2(\alpha_3)]/B_1^2(\alpha_3), \end{cases} \tag{2.2}$$

$$M_1^{-1}: \begin{cases} \alpha_1(s, t) = (1-s)[1 - 1/(\sqrt{1+t^2} + 1 - t)], \\ \alpha_2(s, t) = s[1 - 1/(\sqrt{1+t^2} + 1 - t)], \\ \alpha_3(s, t) = 1/(\sqrt{1+t^2} + 1 - t). \end{cases} \tag{2.3}$$

Example 2.6.

$$M_3: \begin{cases} s = v, \\ t = [B_2^2(u) - B_0^2(u)]/B_1^2(u), \end{cases} \tag{2.4}$$

$$M_3^{-1}: \begin{cases} u(s,t) = 1/(\sqrt{1+t^2}+1-t), \\ v(s,t) = s. \end{cases} \tag{2.5}$$

Example 2.7.

$$M_4: \begin{cases} s = [u(1-v)]/[u(1-v)+(1-u)v], \\ t = (1-u-v)/[u(1-v)+(1-u)v], \end{cases} \tag{2.6}$$

$$M_4^{-1}: \begin{cases} u(s,t) = 2s/[t+2s+\sqrt{t^2+4s(1-s)}], \\ v(s,t) = 2(1-s)/[t+2(1-s)+\sqrt{t^2+4s(1-s)}], \end{cases} \tag{2.7}$$

where (u, v) are defined by the limit when $s = 0$ or $s = 1$.

3. Approximating curve classes–reduced form algebraic curves

The reduced form algebraic curves are a special form of regular algebraic curve segments discussed in (Xu et al., 2000). In this special form, we take most of the BB-form coefficients to be zero and arrange the nonzero coefficients on three parallel lines (see the dots in Figs. 5–7). Here, we define three reduced form curve classes: HT_m (Horizontal form on Triangle), VS_m(Vertical form on Square) and DS_m (Diagonal form on Square), where m is a parameter relating to the degree of the BB-form.

A. Horizontal form HT_m. This class is a subset of $D_1([p_1p_2p_3], p_3, [p_1p_2])$-regular curves (see Theorem 4.1 of (Xu et al., 2000)) defined by:

$$HT_m = \left\{ F(\alpha) = 0: \ 0 < \alpha_3 < 1, \ F = \frac{2}{m+1} \sum_{i=0}^{m} \beta_i B_{m-i,i,1}^{m+1}(\alpha) \right.$$

$$- \frac{2}{m(m+1)} \sum_{i=0}^{m-1} w_i B_{m-1-i,i,2}^{m+1}(\alpha) + \sum_{i=0}^{m+1} w_i' B_{m+1-i,i,0}^{m+1}(\alpha),$$

$$\left. \sum_{i=0}^{m-1} w_i B_i^{m-1} > 0; \ w_i' = \sum_{j=i-2}^{i} \frac{C_{n-1}^{j} C_2^{i-j} w_j}{C_{n+1}^{i}} \right\},$$

where w_i' are given by degree elevation formula so that $\sum_{i=0}^{m-1} w_i B_i^{m-1} = \sum_{i=0}^{m+1} w_i' B_i^{m+1}$. The functions F in defining HT_m have three summations. The first, that has degree m on α_1 and α_2, relates to the terms on the middle line of Fig. 5. The second, that has degree $m-1$ on α_1 and α_2, relates to the terms on the top line. The third, that has degree $m+1$ on α_1 and α_2, relates to the terms on the third line. The degree elevation will make the third have degree $m-1$ too, so that we could transform the equation $F(\alpha) = 0$ into rational form (5.1). The curves in HT_m are between p_3 and $[p_1p_2]$ and away from them (see the curve in Fig. 5).
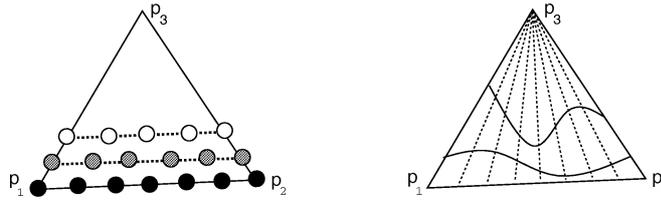
Fig. 5. Horizontal form: Dots represent non-zero Bézier coefficients. Function on the line with real/empty dots is positive/negative. The corresponding curves are between p_3 and $[p_1 p_2]$.
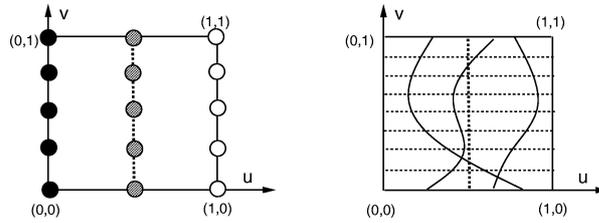
Fig. 6. Vertical form: Dots represent non-zero Bézier coefficients. Function on the line with real/empty dots is positive/negative. The corresponding curves are between the line $u = 0$ and the line $u = 1$.
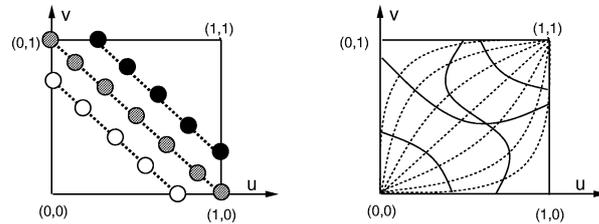
Fig. 7. Diagonal form: Dots represent non-zero Bézier coefficients. Function on the line with real/empty dots is positive/negative. The corresponding curves are between the point $(0, 0)$ and point $(1, 1)$.

B. Vertical form VS_m. This class of curves is a subset of $D_3([p_1 p_2 p_3 p_4], [p_1 p_2], [p_3 p_4])$-regular curves (see Theorem 4.5 of (Xu et al., 2000)) defined by:

$$VS_m = \left\{ F(u, v) = 0: \ 0 < u < 1, \ 0 \leqslant v \leqslant 1, \ F(u, v) = B_1^2(u) \sum_{i=0}^{m} \beta_i B_i^m(v) \right.$$

$$\left. + \left[B_0^2(u) - B_2^2(u) \right] \sum_{i=0}^{m-1} w_i B_i^{m-1}(v), \ \sum_{i=0}^{m-1} w_i B_i^{m-1} > 0 \right\}.$$

The curves in this set are between $[p_1 p_2]$ and $[p_3 p_4]$ and away from them (see the curve in Fig. 6).

C. Diagonal form DS_m. This class of curves is a subset of $D_4([p_1 p_2 p_3 p_4],\ p_1,\ p_4)$-regular curves (see Theorem 4.8 of (Xu et al., 2000)) defined by:

$$DS_m = \left\{ F(u,v) = 0: \ (u,v) \in [0,1] \times [0,1] \setminus \{(0,0),(1,1)\}, \right.$$

$$F(u,v) = \sum_{i=0}^{m} \gamma_i \beta_i B_i^m(u) B_{m-i}^m(v) - \sum_{i=0}^{m-1} \delta_i w_i B_i^m(u) B_{m-i-1}^m(v)$$

$$\left. + \sum_{i=0}^{m-1} \eta_i w_i B_{i+1}^m(u) B_{m-i}^m(v), \ \sum_{i=0}^{m-1} w_i B_i^{m-1} > 0 \right\},$$

where $\gamma_i = 1/C_m^i$, $\delta_i = C_{m-1}^i/(C_m^{i+1} C_m^i)$, $\eta_i = C_{m-1}^i/(C_m^{i+1} C_m^{m-i})$ are constants. These constants are chosen so that $F(u,v) = 0$ could be transformed into the form (5.1). The curves in DS_m are between p_1 and p_4 and away from these two points (see the curves in Fig. 7).

In the definition of these curve families, the parameters β_i are free but $\sum_{i=0}^{m-1} w_i$ $B_i^{m-1}(s) > 0$ on $[0,1]$. If we cannot guarantee the validity of this condition in the interpolation and approximation problems, we can simply take $w_i = 1$. However, we then lose $m-1$ degrees of freedom.

4. Approximated curve classes

We have defined three classes of regular algebraic curves in Section 3. Each class has different features which make them suitable for approximating different curve classes. In this section, we define these curve classes.

Let $D(R, R_1, R_2)$ *be a given discriminating family. Then the approximated curve class corresponding to* $D(R, R_1, R_2)$ *is the collection of curves* $g(x,y) = 0$ *that intersect each curve in* $D(R, R_1, R_2)$ *once and only once.* Here we assume $g(x,y)$ is C^1 continuous in the region R. Denote this class by $A(R, R_1, R_2)$. The curve in $A(R, R_1, R_2)$ has the following features:

(a) It is smooth in $R \setminus \{R_1, R_2\}$.
(b) It is away from R_1 and R_2.
(c) If a point p in $R \setminus \{R_1, R_2\}$ is on the curve $g(x,y) = 0 \in A(R, R_1, R_2)$ and the curve $d(x,y) = 0 \in D(R, R_1, R_2)$, then the two curves are not tangent at p.

For a given curve $g(x,y) = 0$ in $A(R, R_1, R_2)$ there may be infinitely many functions $g(x,y)$ that have the same zero contour. That is, they represent one element in $A(R, R_1, R_2)$. We shall show that all these functions can be represented by one single function $f \in C^1[0,1]$. We say $g(x,y)$ and $f(s)$ are equivalent in defining the curve. Here $C^1[0,1]$ is the set of all C^1 continuous functions on $[0,1]$.

Theorem 4.1. *Let* $D(R, R_1, R_2) = \{s = \alpha(x,y): \ s \in [0,1]\}$ *be a discriminating family, and* $T(R, R_1', R_2') = \{t = \beta(x,y): \ t \in (-\infty, \infty)\}$ *be a transversal family of* $D(R, R_1, R_2)$.
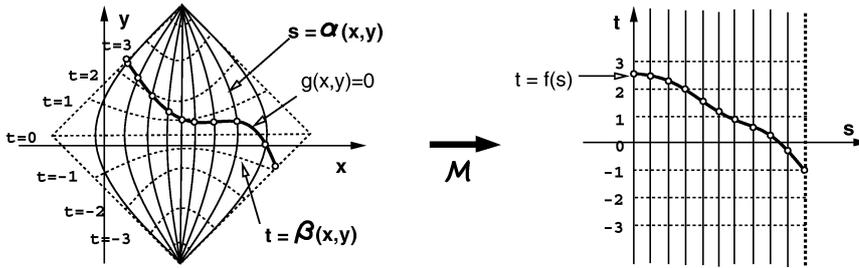
Fig. 8. A curve $g(x, y) = 0$ in $A(R, R_1, R_2)$ is mapped into a function $f(s)$ in $C^1[0, 1]$. The small circles on the left (or right) figure are the intersection points of the curve $g(x, y) = 0$ (or $t = f(s)$) with the the curves $s_i = \alpha(x, y)$ (or the lines $s = s_i$).

Then $g(x, y) = 0 \in A(R, R_1, R_2)$ if and only if there exists a unique $f \in C^1[0, 1]$ such that the curve $g(x, y) = 0$ is equivalently defined by $\beta(x, y) - f(\alpha(x, y)) = 0$.

Proof. Let $g(x, y) = 0 \in A(R, R_1, R_2)$. Then by the definition of $A(R, R_1, R_2)$, we know that for any $s \in [0, 1]$, the curve $s = \alpha(x, y) \in D(R, R_1, R_2)$ will intersect with $g(x, y) = 0$ only once in $R \setminus \{R_1, R_2\}$. Let $(x(s), y(s))^T$ be the intersection point. Then by the implicit function theorem, $x(s)$ and $y(s)$ are C^1 continuous functions of s and $(x(s), y(s))^T \notin R_1 \cup R_2$. Let $f(s) = \beta(x(s), y(s))$. Then $f \in C^1[0, 1]$. Now for this f, $\beta(x, y) - f(\alpha(x, y)) = 0$ defines the curve $g(x, y) = 0$ since $(x(s), y(s))^T$ satisfies the equation. It is easy to see that this f is unique. Hence the necessary part of the theorem is proved.

Let $f \in C^1[0, 1]$, then $g(x, y) = \beta(x, y) - f(\alpha(x, y))$ is clearly C^1 continuous on $R \setminus \{R_1, R_2\}$. For any given $s \in [0, 1]$ the curve $s = \alpha(x, y)$ intersects with $g(x, y) = 0$ at a point $(x, y)^T$ that satisfies (2.1) with $t = f(s)$. Hence, there exist a unique $(x, y)^T \in R \setminus \{R_1, R_2\}$ that satisfies (2.1). Since

$$\nabla g = \nabla \beta - \nabla \alpha f' \tag{4.1}$$

then by the definition of the transversal family, we know that $\nabla g \neq 0$. That is, the curve $g(x, y) = 0$ is smooth. Since $\nabla \alpha$ and $\nabla \beta$ are linearly independent, Eq. (4.1) also implies that ∇g and $\nabla \alpha$ are linearly independent. That is, the intersection is only once. Therefore $g(x, y) = 0 \in A(R, R_1, R_2)$. $\square$

Theorem 4.1 gives an invertible mapping from $A(R, R_1, R_2)$ to $C^1[0, 1]$ that maps a zero contour $g(x, y) = 0 \in A(R, R_1, R_2)$ to a function $f \in C^1[0, 1]$. We denote this map by $\mathcal{M}(\{g = 0\}) = f$ (see Fig. 8).

For the three cases considered, we denote $A(R, R_1, R_2)$ as HT, VS, and DS, respectively, in contrast with the approximating families HT_m, VS_m, and DS_m.

5. Interpolation and approximation

We consider now the main problem: interpolate or approximate the curves in $A(R, R_1, R_2)$ by the reduced form algebraic curves. By interpolation, we mean that we are given a

set of data points on a curve and possibly derivatives, and we wish to construct another curve in the reduced form class to contain these points and match derivatives. For the approximation problem, we are given a smooth curve in $A(R, R_1, R_2)$ and then wish to construct an algebraic curve to approximate it. The interpolation and approximation for a given curve $g(x, y) = 0 \in A(R, R_1, R_2)$ will be realized by the following steps:

(1) Find $f \in C^1[0, 1]$, such that $\mathcal{M}(\{g = 0\}) = f$.
(2) Determine the interpolant or approximant $\tilde{f} \in C^1[0, 1]$ of f.
(3) Determine a reduced form curve $\tilde{g}(x, y) = 0 \in A(R, R_1, R_2)$ such that $\{\tilde{g} = 0\} = \mathcal{M}^{-1}(\tilde{f})$. Then $\tilde{g} = 0$ is the approximation of the curve $g = 0$.

Constructing f in the first step is described in Section 5.1. As we shall see in the following, the problem of determining $\tilde{f}$ in the second step will lead to a rational polynomial problem (see Sections 5.2–5.3). That is, determine the coefficients β_i, $i = 0, 1, \ldots, m$, and w_i, $i = 0, 1, \ldots, m - 1$, such that

$$\frac{\sum_{i=0}^{m} \beta_i B_i^m(s)}{\sum_{i=0}^{m-1} w_i B_i^{m-1}(s)} = t(s), \tag{5.1}$$

approximately for a known function $t(s) \in C^1[0, 1]$. After β_i and w_i are determined, computing $\tilde{g}$ in step (3) is obvious, since these β_i and w_i are the required ones in defining the reduced form curves. Now we show how to arrive at problem (5.1) from our reduced form algebraic curve interpolation and approximation problems.

5.1. Convert approximated curves to C^1 functions

Let $g(x, y) = 0 \in A(R, R_1, R_2)$. Then by the definition of $A(R, R_1, R_2)$, we know that for any $s \in [0, 1]$, the curve $s = \alpha(x, y) \in D(R, R_1, R_2)$ intersects with $g(x, y) = 0$ only once in $R \setminus \{R_1, R_2\}$. Let $(x(s), y(s))^T$ be the intersection point. Then by the implicit function theorem, $x(s)$ and $y(s)$ are C^1 continuous functions of s and $(x(s), y(s))^T \notin R_1 \cup R_2$. Let

$$f(s) = \beta(x(s), y(s)). \tag{5.2}$$

Then $f \in C^1[0, 1]$. Now for this f, $\beta(x, y) - f(\alpha(x, y)) = 0$ defines the curve $g(x, y) = 0$ since $(x(s), y(s))^T$ satisfies the equation. It is easy to see that this f is unique.

If $(x(s), y(s))^T$ have closed forms, then (5.2) gives an explicit expression for f. In any other case, function values $f(s)$ could be obtained by the process provided above for $s \in [0, 1]$.

5.2. Convert reduced form curves to rational functions

A. Horizontal form HT_m. Using the first equality of map (2.2), curve $F(\alpha)$ defined in HT_m can be written as $F(\alpha) = (1 - \alpha_3)^{m-1} G(s, t)$ with

$$G(s, t) = B_1^2(\alpha_3) \sum_{i=0}^{m} \beta_i B_i^m(s) - \left[B_2^2(\alpha_3) - B_0^2(\alpha_3) \right] \sum_{i=0}^{m-1} w_i B_i^{m-1}(s). \tag{5.3}$$

Since $0 < \alpha_3 < 1$, the curve $F(\alpha) = 0$ is equivalently defined by $G(s, t) = 0$. By the second equality of map (2.2), $G(s, t) = 0$ can be written in the form (5.1) since $B_1^2(\alpha_3) > 0$ on $(0, 1)$.

B. Vertical form VS_m.   In VS_m, $F(u, v) = G(s, t)$, where $G(s, t)$ is in the same form as (5.3). Hence $G(s, t) = 0$ can be written in the form (5.1) (see (2.4)).

C. Diagonal form DS_m.   By map (2.6), we can write $F(u, v)$ defined in DS_m as $F(u, v) = [u(1-v) + (1-u)v]^{m-1} G(s, t)$, where

$$G(s, t) = [u(1-v) + (1-u)v] \sum_{i=0}^{m} \beta_i B_i^m(s) - (1-u-v) \sum_{i=0}^{m-1} w_i B_i^{m-1}(s) \quad (5.4)$$

and u and v are defined by (2.7). Since $(u, v) \neq (0, 0)$, $(1, 1)$, $u(1-v) + (1-u)v > 0$. Hence $F(u, v) = 0$ if and only if $G(s, t) = 0$. By map (2.6), $G(s, t) = 0$ can be written as (5.1).

5.3. Rational interpolation and approximation

Having arrived at the problem of determining coefficients β_i and w_i such that (5.1) holds, where $(s, t)^{\mathrm{T}}$ is related to $(x, y)^{\mathrm{T}}$ by the known map (2.1), we now show how the interpolation and approximation problems can be solved as classical rational ones.

A. Hermite interpolation
Suppose we are given some points with derivatives in the domain considered:

$$\left(x_j, \ y(x_j), \ y^{(1)}(x_j), \ldots, y^{(k_j)}(y_j) \right), \quad j = 0, 1, \ldots, n, \quad (5.5)$$

or $(y_j, \ x(y_j), \ x^{(1)}(y_j), \ldots, x^{(k_j)}(y_j))$, $j = 0, 1, \ldots, n$, such that $s_j = \alpha(x_j, \ y_j)$ are distinct and

$$\sum_{j=0}^{n} k_j + 1 \geqslant 2m \quad \text{(or } m+1 \text{ if } w_i \text{ are fixed)}. \quad (5.6)$$

From the discussion above, s_j distinct means that (x_j, y_j) is separable by the corresponding discriminating family. That is, each curve in the discriminating family contains at most one (x_j, y_j). Then using (2.1) we can compute the points $\{s_j, t(s_j), t^{(1)}(s_j), \ldots, t^{(k_j)}(s_j)\}$, $j = 0, 1, \ldots, n$. For example, if (5.5) is given, then

$$s_j = \alpha\left(x_j, y(x_j)\right),$$

$$t(s_j) = \beta\left(x_j, y(x_j)\right),$$

$$t'(s_j) = \left[\frac{\partial \beta}{\partial x} + \frac{\partial \beta}{\partial y} y'(x_j)\right] \frac{\partial x}{\partial s},$$

$$t''(s_j) = \left[\frac{\partial^2 \beta}{\partial x^2} + 2 \frac{\partial^2 \beta}{\partial x \partial y} y'(x_j) + \frac{\partial^2 \beta}{\partial y^2} \left(y'(x_j)\right)^2 + \frac{\partial \beta}{\partial y} y''(x_j)\right] \frac{\partial x}{\partial s}$$

$$+ \left[\frac{\partial \beta}{\partial x} + \frac{\partial \beta}{\partial y} y'(x_j)\right] \frac{\partial^2 x}{\partial s^2}, \quad \cdots$$

where $\frac{\partial x}{\partial s}$, $\frac{\partial^2 x}{\partial s^2}$, ... are determined by differentiating $s = \alpha(x, y)$. That is

$$1 = \left[\frac{\partial \alpha}{\partial x} + \frac{\partial \alpha}{\partial y} y'(x_j) \right] \frac{\partial x}{\partial s},$$

$$0 = \left[\frac{\partial^2 \alpha}{\partial x^2} + 2 \frac{\partial^2 \alpha}{\partial x \partial y} y'(x_j) + \frac{\partial^2 \alpha}{\partial y^2} \left(y'(x_j) \right)^2 + \frac{\partial \alpha}{\partial y} y''(x_j) \right] \frac{\partial x}{\partial s}$$

$$+ \left[\frac{\partial \alpha}{\partial x} + \frac{\partial \alpha}{\partial y} y'(x_j) \right] \frac{\partial^2 x}{\partial s^2}, \cdots.$$

It is clear that $\frac{\partial^k x}{\partial s^k}$ is well defined if $\frac{\partial \alpha}{\partial x} + \frac{\partial \alpha}{\partial y} y'(x_j) \neq 0$. The geometric meaning of this requirement is that the curve constructed is not tangent with the curve in the discriminating family at the interpolating points. Differentiating (5.1) about s and then evaluating it at s_j, we get a linear system of equations with β_i and w_i as unknowns. It is a classical Hermite interpolation problem. Under condition (5.6) it can be solved in the least square or Chebyshev sense. If we interpolate two points, the highest order of smoothness we can achieve is C^{m-1}.

If the points (5.5) come from a curve that is in the corresponding approximated curve class discussed in Section 4, then the interpolation error formula for the function $t(s)$ can be used. Of course, well behaved interpolation schemes, such as Fejér interpolation, or interpolation on Chebyshev points, can be used as well. In summary, there are plenty of methods and results for rational or polynomial interpolation can be applied here.

B. Approximation

Suppose we are given a curve in the corresponding class defined in Section 4. In each case, map (2.1) leads us to a function $t = t(s) \in C^1[0, 1]$ (see Section 5.1), and the approximation problem is transformed to the problem of determining the coefficients, such that (5.1) holds approximately. In any case as discussed before, $t(s)$ is always a well defined C^1 continuous function. So we are led to the classical rational approximation problem, and the related methods and results can be used. Hence we have proved the following theorem:

Theorem 5.1. *The interpolation and approximation problems of the curves in the approximated curve classes by the curves in the reduced form curves classes always have a unique solution.*

5.4. Parametric representation of the solutions

After the coefficients β_i, $i = 0, 1, \ldots, m$, and w_i, $i = 0, 1, \ldots, m - 1$, are determined by solving problem (5.1), the implicit form solution is obtained for our proposed problem. In this sub-section, we use the map M^{-1} to obtain a parametric form solution. This parametric form is important for efficiently evaluating the curve at several points for display. Now let $t_m(s) = \sum_{i=0}^{m} \beta_i B_i^m(s) / \sum_{i=0}^{m-1} w_i B_i^{m-1}(s)$.

A. Horizontal form HT_m. From (5.3), it is easy to find that

$$\alpha_i(s) = \alpha_i \left(s, t_m(s) \right), \quad i = 0, 1, 2, \tag{5.7}$$

where $\alpha_i(s, t)$ is explicitly defined by (2.3).

B. Vertical form VS_m.   Since $G(s, t)$ is in the same form of (5.3), we have

$$\begin{cases} u(s) = 1/\left[\sqrt{1 + t_m(s)^2} + 1 - t_m(s)\right], \\ v(s) = s. \end{cases} \tag{5.8}$$

C. Diagonal form DS_m.   From $G(s, t) = 0$, where $G(s, t)$ is defined by (5.4), we can obtain

$$\begin{cases} u(s) = u(s, t_m(s)), \\ v(s) = v(s, t_m(s)), \end{cases} \tag{5.9}$$

where $(u(s, t), v(s, t))^{\mathrm{T}}$ is defined by (2.7).

5.5. Examples

To illustrate that the error of the interpolant or approximant by D-regular algebraic curves are controllable, we give the following example.

Example 5.1. Consider the curve $(x, f(x))^{\mathrm{T}}$ with $x \in [-1, 1]$ and $f(x) = 1/(1 + 25x^2)$. We interpolate the curve at the points $(x_i, f(x_i))^{\mathrm{T}}$ with $x_i = -1 + (2i/m)$, $i = 0, 1, \ldots, m$, and $m = 2, 4, \ldots, 16$. The function $f(x)$ in this example is notorious in polynomial interpolation in that the interpolation error goes to infinity when the degree m of the interpolation polynomial goes to infinity. Table 1 gives the maximal errors of polynomial interpolants (the second column), D_1-regular curve interpolants (the third column), D_3-regular curve interpolants (the fourth column) and D_4-regular curve interpolants (the fifth column), respectively. For the D_1-regular curve, the vertices of the triangle, say $[p_1 p_2 p_3]$, are taken to be $p_1 = (-1, 1/26)^{\mathrm{T}}$, $p_2 = (1, 1/26)^{\mathrm{T}}$ and $p_3 = (0, 1.2)^{\mathrm{T}}$. For the D_3-regular curve, the vertices of the quadrilateral, say $[p_1 p_2 p_3 p_4]$, are taken to be $p_1 = (1, 0)^{\mathrm{T}}$, $p_2 = (-1, 0)^{\mathrm{T}}$, $p_3 = (1, 1 + 1/26)^{\mathrm{T}}$ and $p_4 = (-1, 1 + 1/26)^{\mathrm{T}}$. For the D_4-regular curve, the vertices of the quadrilateral, say $[p_1 p_2 p_3 p_4]$, are taken to be $p_1 = (0, -1 + 1/26)^{\mathrm{T}}$, $p_2 = (-1, 1/26)^{\mathrm{T}}$, $p_3 = (1, 1/26)^{\mathrm{T}}$ and $p_4 = (0, 1 + 1/26)^{\mathrm{T}}$. To be fair in the comparison, all the w_i in defining the D-regular curves are taken to be one so that we use the same

Table 1
The maximal errors of the interpolants

m	Polynomial	D_1-regular	D_3-regular	D_4-regular
2	0.452766	0.042876	0.538708	0.299358
4	0.434777	0.017383	0.296371	0.595020
6	0.612582	0.013753	0.775354	0.652474
8	1.041586	0.015925	0.531750	0.527474
10	1.336609	0.014871	0.910885	0.233677
12	3.659931	0.010442	0.735780	0.044549
14	6.309690	0.015417	0.926017	0.027907
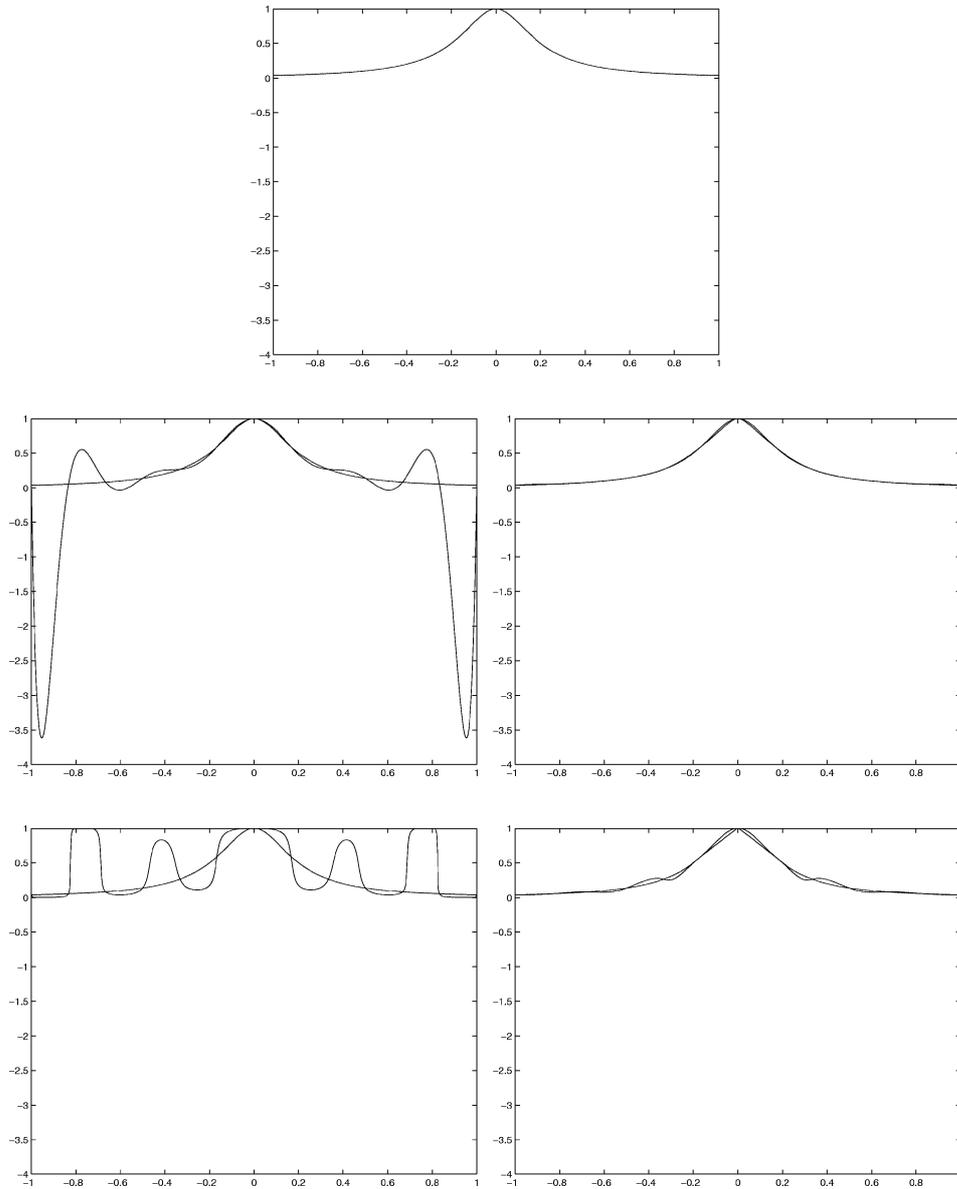
Fig. 9. The top box shows the original curve. Each of the other boxes shows the original curve and an interpolation curve. Middle left: The polynomial interpolation curve; Middle right: The D_1-regular interpolation curve; Bottom left: D_3-regular interpolation curve; Bottom right: D_4-regular interpolation curve.

number of degrees of freedom and interpolate the same set of points. The maximal error is taken to be the Hausdorff distance of two curves defined by

$$h(f, g) = \max\Big\{ \sup_{p \in f=0} \inf_{q \in g=0} \|p - q\|, \ \sup_{q \in g=0} \inf_{p \in f=0} \|p - q\| \Big\}$$

for the curves $f = 0$ and $g = 0$.

It should be noted from Table 1 that the errors of the polynomial interpolation increase rapidly, while the errors of the D-regular interpolants are always limited. Especially, note that the maximal errors of the D_1-regular curve are significantly reduced. The maximal errors of other D-regular curve are also reduced when m becomes larger. In Fig. 9, we plot the original curve and interpolation curves for $m = 12$. The figure shows that D_1 and D_4 regular curves yield very good approximations of the given curve. The D_3 regular curve does not look good for this example, but the error is much smaller than its polynomial counterpart.

Though the errors are indeed controllable in this example using D-regular algebraic curves, we are not claiming that the errors of reduced form algebraic curves are always smaller than their polynomial counterpart. This uncertainty is due to the face that each reduced form curve class has its own favorite curves for approximation. Consider a rational function $t = R(s)$ in the st-plane. Under the transform $\mathcal{M}^{-1}$, it maps to a curve $A(x, y) = 0$ in the xy-plane in each of the three cases. If we take such a curve as the approximated one, then its reduced form approximation will of course be exact. However, the curve $A(x, y) = 0$ is no longer a rational function in the xy-plane in general. Its rational approximation will have error. Therefore, it is quite easy to construct examples, such that

Table 2
The maximal errors of the rational interpolants

m, n	Rational	D_1-regular	D_3-regular	D_4-regular
2,0	0.195382	0.301863	0.298608	0.110803
4,0	0.081493	0.205984	0.169827	0.072066
6,0	0.303969	0.011034	0.008595	0.038220
8,0	0.274009	0.001622	0.002572	0.012411
2,2	0.092865	0.093109	0.091161	0.077113
4,2	singular	0.017835	0.014409	singular
6,2	singular	0.001159	0.002347	0.013134
8,2	0.101666	0.000530	0.000971	singular
2,4	0.016656	singular	singular	0.026142
4,4	0.004224	0.004275	0.003983	0.009034
6,4	0.002841	0.000533	0.001094	0.011416
8,4	0.001601	0.000531	0.000270	0.006222

their reduced form approximation have smaller errors. To the contrary, examples, where the rational approximation yields lower errors, are also easy to provide. The next example (Example 5.2) shows this uncertainty, where there are several cases in which the errors of D_1, D_3 and D_4 regular curves are higher than the rational curves, and there are several other cases in which the errors of D_1, D_3 and D_4 regular curves are lower.

Example 5.2. Let $f(x) = 1/1 + 25|x^5|$. We use degree (m, n) rational function to interpolate the function on uniform knots on $[-1, 1]$. The function $t(s)$ that defined the reduced form algebraic curves are also (m, n) rational function. Table 2 gives the maximal Hausdorff distances of the original curve and the interpolants. The entry "singular" in the table means the rational function determined by the interpolation condition has poles.

6. Conclusions

Corresponding to the discriminating family introduced in (Xu et al., 2000), we have introduced a notion of a transversal family. Using this pair of curve families, we have transformed the reduced form algebraic curves into rational functions. Hence, the problems of interpolation and approximation with reduced form algebraic curves in the xy-plane are solved by first converting the problems into classical rational form in the st-plane, and then solving the rational problems there and transforming the results back to the original plane. The curves so constructed are non-singular and without discontinuities. Furthermore, the explicit formulas are given for evaluating the curves efficiently. Unlike the polynomial rational functions, the curves constructed are always located in the triangular or quadrilateral domain. Hence, the error of the interpolant or approximant by the algebraic curve is controllable.

In this paper, we have not addressed the problem of constructing piecewise reduced form algebraic curves, because this is not difficult to achieve with a local interpolation approach. To get a G^k smooth piecewise algebraic curve that interpolates a sequence of join points with C^k data attached at each point, we can simply construct (see (Bajaj and Xu, 2000)) a triangle or quadrilateral for each pair of adjacent join points and interpolate the C^k data at the two points. The highest order of smoothness we can achieve with a reduced form algebraic curve is $m - 1$, since the curve has $2m$ degrees of freedom.

References

Bajaj, C. and Xu, G. (1996), Data fitting with cubic A-splines, in: Gigante, M. and Kunii, T. (Eds.), Proc. of Computer Graphics International, CGI'94, Melbourne, Australia, World Scientific Publishing, 338–346.

Bajaj, C. and Xu, G. (1999), A-splines: Local interpolation and approximation using G^k-continuous piecewise real algebraic curves, Computer Aided Geom. Design 16, 557–578.

Bajaj, C. and Xu, G. (2000), Regular algebraic curve segments (III)—Applications in data fitting, Computer Aided Geom. Design, submitted.

Paluszny, M. and Patterson, R.R. (1992), Curvature continuous cubic algebraic splines, in: Proc. of SPIE, Curves & Surfaces in Computer Vision and Graphics III, Vol. 1830, 48–57.

Paluszny, M. and Patterson, R.R. (1993), A family of tangent continuous algebraic splines, ACM Transaction on Graphics 12 (3), 209–232.

Paluszny, M. and Patterson, R.R. (1998), Geometric control of G^2-cubic A-splines, Computer Aided Geom. Design 13, 261–287.

Sederberg, T., Anderson, D. and Goldman, R. (1985), Implicitization, inversion, and intersection of planar rational cubic curves, Computer Vision, Graphics and Image Processing 31, 89–102.

Xu, G., Bajaj, C. and Xue, W. (2000), Regular algebraic curve segments (I)—Definitions and characteristics, Computer Aided Geom. Design 17, 485–501.