Numerically Stable Implicitization of Cubic Curves

John D. Hobby

Abstract

We give efficient, numerically-stable techniques for converting polynomial and rational cubic curves to implicit form. We achieve numerical stability by working in a rotated coordinate system and using carefully chosen expressions for the coefficients that appear in the implicit form. This is more practical than previously known methods which can be numerically unstable unless all computations are done in exact rational arithmetic.

Categories and Subject Descriptors: I.3.5 [**Computer Graphics**]: Computational Geometry and Object Modeling—Curve, surface, solid, and object representations

General Terms: Algorithms, Reliability

Additional Key Words and Phrases: Cubic curves, implicitization, numerical stability

Numerically Stable Implicitization of Cubic Curves

John D. Hobby

1. Introduction

The problem of converting parametrically defined curves to implicit form has been studied extensively by Sederberg [6, 8, 9, 10]. As Sederberg explains, it is sometimes convenient to have an implicit equation $F(x, y) = 0$ where F is a bivariate polynomial. The implicit form is favored for applications that require testing whether given points lie on the curve or on one side or the other. Sederberg [11] gives applications to curve intersection; Hobby [4] gives applications to rasterization.

Curves that can be expressed in implicit form are called algebraic curves. (See Semple and Roth [13]). Algebraic curves include polynomial curves $x = f(t)$, $y = g(t)$ where f and g are polynomials, and rational curves $x = f(t)/h(t)$, $y = g(t)/h(t)$ where f, g and h are polynomials. Polynomial curves are commonly pieced together and represented as B-splines, Beziér curves, or beta-splines. Commonly used rational curves are conic sections and curves obtained by taking projective transformations of polynomial curves. Of course any polynomial curve can be thought of as a rational curve where $h(t)$ is constant.

The implicit form for a rational curve $x = f(t)/h(t)$, $y = g(t)/h(t)$ is the resultant of the polynomials $f - xh$ and $g - yh$. (The resultant of two polynomials is an expression that must be zero in order for the polynomials to have a common root. [10]) Since the resultant can be expressed as the determinant of a matrix whose nonzero entries are coefficients of the polynomials, we can find the implicit form by evaluating the determinant of a matrix whose entries are linear polynomials in x or y.

Cubic curves are often used in practice, and special properties of cubic curves allow an implicitization process that is substantially more efficient than evaluating the resultant directly. [9] A rational cubic curve can cross itself at most once, and there is a simple way to find the implicit form by first shifting the coordinate system so that the crossing point lies at the origin. The same technique works even when there is no crossing if the concept of a "crossing point" is suitably generalized.

A significant disadvantage of the above methods is that they do not take numerical stability into account. Rather, it is generally assumed that all computations are to be carried out in exact rational arithmetic. While rational arithmetic has some important advantages, it tends to produce large numbers that are expensive manipulate. For instance, if the coefficients of the cubic polynomials f, g and h are fixed-point numbers n-bits long, the resultant is a six-by-six determinant and the coefficients in the implicit form $F(x, y) = 0$ each require about $6n$ bits. In other words it may be necessary to deal with large integers whose length is up six times the normal word size.

The purpose of this paper is to make implicitization more practical by allowing much or all of the work to be done in floating-point, while still producing an implicit form that represents a curve close to the desired one. In other words, the desired curve corresponds to an interval $0 \leq t \leq 1$ in parameter space and we want good error bounds for the corresponding portion of the curve represented by the computed implicit form. A major difficulty is that small changes in the desired curve can result in large changes in the implicit form. Thus we have the difficult task ensuring that $F(x, y) = 0$ is close to the desired curve even though the coefficients of $F(x, y)$ cannot be computed accurately.

To see the difficulty in computing coefficients, consider the polynomial cubic

$$(x, y) = (1-t)^3(0,0) + 3t(1-t)^2(263,110) + 3t^2(1-t)(427,205) + t^3(519,285)$$

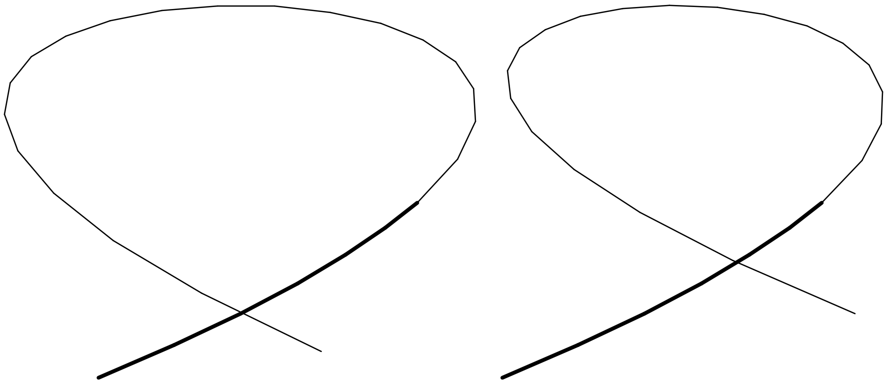
shown in Figure 1a. It has a crossing at $(231, 105)$ and an implicit form

$$-19683y^3 + 16041645y^2 - 2460375x^2 - 2717730225y + 1136693250x = 0.$$

Changing the final control point $(519, 285)$ to $(520, 285)$ has little effect on the portion of the curve where $0 \le t \le 1$, but as shown in Figure 1b, it moves the crossing to a point near $(382, 188)$. The resulting implicit form also has major changes:

$$-21952y^3 + 21368340y^2 - 1247400xy - 2551500x^2 - 5224205700y + 2185029000x = 0.$$

Thus it is an ill-conditioned problem to compute the coefficients of powers of x and y in $F(x, y) = 0$.



(a) (b)

Figure 1: (a) The curve $(1-t)^3(0,0) + 3t(1-t)^2(263,110) + 3t^2(1-t)(427,205) + t^3(519,285)$ with the portion where $0 \le t \le 1$ shown in bold; (b) the curve $(1-t)^3(0,0) + 3t(1-t)^2(263,110) + 3t^2(1-t)(427,205) + t^3(520,285)$ with similar behavior on $0 \le t \le 1$ but a significantly different implicit form.

Even if we could compute the coefficients of $F(x, y) = 0$ with good relative accuracy there is no guarantee that the computed implicit form represents a curve close to the desired one. The problem arises when the parametric curve is very close to a straight line such as

$$(x, y) = (1-t)^3(0,0) + 3t(1-t)^2(10,10) + 3t^2(1-t)(18,18) + t^3(23,23). \tag{1}$$

(The choice of a precisely straight "curve" simplifies the example—the same issues arise in practice for similar curves that are not quite straight). In this example, we can see by inspection that the curve eventually doubles back on itself, causing the implicit form to have triple roots on the line $y = x$. In other words, the implicit form is $(y - x)^3 = 0$.

To observe the sensitivity to small changes in the implicit form, consider the effect of multiplying the x^3 term by $1 + \delta$ for some small δ. This changes the implicit form to $(y - x)^3 - \delta x^3 = 0$, so that the zeros now lie on the line $y = (1 + \sqrt[3]{\delta}\,)x$. Thus the deviation from the $(0, 0)$ to $(23, 23)$ segment is essentially $(23/\sqrt{2}\,)\sqrt[3]{\delta}$ rather than on the order of δ. The strategy for avoiding this cube root is to rotate the coordinate system 45° so that the line falls on one of the coordinate axes and the implicit form is just $x^3 = 0$ or $y^3 = 0$.

How do existing implicitization algorithms deal with the difficulty in computing the coefficients of $F(x, y) = 0$ and the need for a rotated coordinate system? The resultant method involves evaluating the determinant of a matrix whose entries are polynomials in x and y. The ill-conditioned nature of the coefficient computations leads to large errors in the implicit form. Such errors do not necessarily distort the curve $F(x, y) = 0$ a lot, but in this case they often do. For instance, using MAPLE [1] to implement the resultant method on the curve in Figure 1 with five-digit floating point arithmetic leads to an implicit curve $F(x, y) = 0$ that never gets within 100 units of the point $(519, 285)$. The maximum error is reduced to 12 units when seven digits are used and 0.48 units with nine digits.

Evaluating the implicitization formulas from [9] in floating point yields somewhat better results, but the maximum error for the above example is still 3.6 units with five digits and 0.96 units with seven digits. This method also suffers from the problem with nearly straight curves in unrotated coordinates.

One implicitization method that does use coordinate transformations for nearly straight curves is due to Sederberg [7]. The idea is to use the parametric form to generate points on the curve in a barycentric coordinate system and use these to define a system of linear equations for the coefficients that appear in the implicit form. Sederberg doesn't discuss the numerical stability of this process, but it is not hard to get bounds on the magnitude of $F(x, y)$ at the generated points. These do not readily yield bounds on the error in the implicit curve $F(x, y) = 0$, especially when there is a crossing point where the derivatives of F approach zero. In the example of the curve from Figure 1, the error in the implicit form is 1.3 units for five-digit floating point and 0.06 units for seven digit floating point.

The main cause for loss of precision in the above example is the flatness of $F(x, y)$ near crossing point. The implicitization method described in the next sections uses special properties of cubic curves to deal with this problem, resulting in a maximum error in the implicit form of 0.00004 units for the above example when using floating point of roughly seven digit precision. Other numerical results are given in Table 1. The curves used in the table are intended to be difficult examples that cover most of the relevant special cases.

parametric form	λ	error	comments
$(-0.1326, 198.8)t_1 + (0.1326, 397.7)t_2 + (0.7954, 596.5)t_3$	1.0	4.6×10^{-8}	double pt. at ∞
$-(181.5, 69.47)t_1 - (365.7, 144.7)t_2 - (552.1, 225.6)t_3$	1.0	0.000065	double pt. far away
$\dfrac{(-5.936, 59.36)t_1 + (11.87, 118.7)t_2 + (17.81, 178.1)t_3}{0.3t_0 + 13107.2t_1 + 13107.2t_2 + 0.3t_3}$	43690.7	0.000007	large λ
$\dfrac{(80.2, -262.1)t_1 + (-320.8, 741.7)t_2 + (899.1, -1573)t_3}{37.21t_0 - 28.67t_1 + 3.05t_2 + 54.29t_3}$	13844.2	0.026674	large λ
$\dfrac{(84.56, 53.7)t_1 + (64.43, 88.56)t_2 + (-60.38, 104.6)t_3}{t_0 + 0.403315t_1 + 0.403315t_2 + t_3}$	1.8	0.000013	arc of an ellipse
$\dfrac{(95.44, 71.96)t_1 - (42.5, 32.02)t_2 + (35.53, 26.77)t_3}{t_0 + 1.59433t_1 + 0.280381t_2 + t_3}$	2.0	0.000027	highly flattened loop
$(263, 110)t_1 + (427, 205)t_2 + (519, 285)t_3$	1.0	0.000039	double pt. on curve

Table 1: Numerical results for computing the implicit form of parametric curves as explained in this paper using 32-bit binary floating point (approximately seven decimal digits). In the parametric form t_0, t_1, t_2 and t_3 stand for $(1-t)^3$, $3t(1-t)^2$, $3t^2(1-t)$ and t^3, respectively. The λ parameter introduced near the beginning of Section 3 is a measure of how "well-behaved" the parametric form is. Note that the second-to-last error figure is computed using the iterative improvement scheme of Section 3.2. Without this, the error would be 0.005196 instead of 0.000027.

Because of the difficulty in achieving guaranteed numerical properties for cubic curves, we make no attempt to extend the analysis to curves of higher order. It may be helpful to look at the background material on the numerical properties of Bézier-Bernstien curves given by Farouki and Rajan [2] and Sederberg and Parry [12].

The overall organization of the paper is as follows: Section 2 derives the implicit form for rational cubic curves and explains how to choose a rotated coordinate system. There is also an alternative form that reduces the error when the crossing point is nearby. Section 3 gives bounds on the round-off error and Section 4 gives some concluding remarks. If the implicit form is needed in unrotated or barycentric coordinates, then appropriate substitutions can be made in $F(x, y)$, using additional precision if necessary.

2. Implicitization of Rational Cubic Curves

Consider the rational cubic curve

$$\big(x(t), y(t)\big) = \left(\frac{f(t)}{h(t)}, \frac{g(t)}{h(t)}\right) \tag{2}$$

obtained by taking a space curve $\big(f(t), g(t), h(t)\big)$ and projecting onto the plane $z = 1$ with the viewpoint at the origin. The purpose of this section is to obtain an implicit form that accurately represents the curve generated as t goes from zero to one.

Such a curve is conveniently represented by the Bézier control points of the space curve. Assuming that the coordinate system has been shifted so that the curve starts at the origin, we have

$$\begin{aligned} 3X_1t(1-t)^2 + 3X_2t^2(1-t) + X_3t^3 &= f(t), \\ 3Y_1t(1-t)^2 + 3Y_2t^2(1-t) + Y_3t^3 &= g(t), \\ Z_0(1-t)^3 + 3Z_1t(1-t)^2 + 3Z_2t^2(1-t) + Z_3t^3 &= h(t). \end{aligned} \tag{3}$$

Note that the same representation works for polynomial cubics if we just set $Z_0 = Z_1 = Z_2 = Z_3 = 1$.

Since it turns out to be desirable not to let $h(t)$ get too small, it is important that a rational reparameterization allows the X_i, Y_i, and Z_i to be changed without effecting the implicit form. (See Patterson [5]). If we select constants α and β and reparameterize by substituting

$$\frac{\alpha t}{\alpha t + \beta(1-t)}$$

for t in (2), factors of $\big(\alpha t + \beta(1-t)\big)^{-3}$ cancel and the net effect is to multiply X_i, Y_i, and Z_i in (3) by $\alpha^i\beta^{3-i}$ for $0 \le i \le 3$. For example, this allows us to force Z_0 and Z_3 to be equal.

We use the same general approach as Sederberg does in [9], but make some important changes to promote numerical stability. The basic idea is to get an expression that takes a point (x, y) on the curve and gives the corresponding t. That expression can then be plugged into anything involving $f(t)$, $g(t)$, or $h(t)$ that is known to be zero when $x = f(t)/h(t)$ and $y = g(t)/h(t)$.

A curve with a crossing point at (x_c, y_c) has a simple relationship between t and the direction of the line from the crossing point to $\big(x(t), y(t)\big)$. This is because $f(t) - x_ch(t)$ and $g(t) - y_ch(t)$ must have two common roots, and this forces the slope

$$\frac{x(t) - x_c}{y(t) - y_c} = \frac{f(t) - x_ch(t)}{g(t) - y_ch(t)} \tag{4}$$

to be the ratio of two linear polynomials in t. In other words, there exists a matrix

$$P = \begin{pmatrix} P_{0x} & P_{1x} \\ P_{0y} & P_{1y} \end{pmatrix}$$

such that

$$\begin{pmatrix} x(t) - x_c & y(t) - y_c \end{pmatrix} \begin{pmatrix} P_{0x} & P_{1x} \\ P_{0y} & P_{1y} \end{pmatrix} \begin{pmatrix} t \\ 1-t \end{pmatrix} = 0. \tag{5}$$

Replacing $x(t)$ by x and $y(t)$ by y in (5) gives the desired relationship between (x, y) and t.

As Sederberg shows in [9], we can always validate the critical assumption that (x_c, y_c) is a crossing point if we suitably generalize the concept of "crossing point." For a generalized crossing point, or more properly *double point*, the two common roots that make (4) a ratio of linear polynomials may be identical (in which case there is a cusp), or they may be complex numbers (when there is no crossing). Additionally, the double point (x_c, y_c) may turn out to be a "point at infinity" (in which case P is singular).

It is possible to avoid dealing with explicit points at infinity by observing that (5) reduces to $(x_c\ y_c)P(0\ 1)^T = 0$ when $t = 0$, hence there is a constant q such that

$$\begin{pmatrix} x_c & y_c \end{pmatrix} \begin{pmatrix} P_{0x} & P_{1x} \\ P_{0y} & P_{1y} \end{pmatrix} \begin{pmatrix} t \\ 1-t \end{pmatrix} = qt. \tag{6}$$

Substituting this into (5) shows that, for a point (x, y) on the curve, the corresponding t satisfies

$$0 = \begin{pmatrix} x & y \end{pmatrix} P \begin{pmatrix} t \\ 1-t \end{pmatrix} - qt = \left(\begin{pmatrix} x & y \end{pmatrix} P - \begin{pmatrix} q & 0 \end{pmatrix} \right) \begin{pmatrix} t \\ 1-t \end{pmatrix}.$$

In other words,

$$\begin{pmatrix} x & y \end{pmatrix} P - \begin{pmatrix} q & 0 \end{pmatrix} \tag{7}$$

is perpendicular to $(t\ 1{-}t)$, hence its two entries give the ratio of $1-t$ to $-t$.

All that remains is to find an expression for q and the elements of P. First substitute (2) and (6) into (5), obtaining

$$\begin{pmatrix} f(t) & g(t) \end{pmatrix} \begin{pmatrix} P_{0x} & P_{1x} \\ P_{0y} & P_{1y} \end{pmatrix} \begin{pmatrix} t \\ 1-t \end{pmatrix} - qth(t) = 0. \tag{8}$$

Since this polynomial equation must hold for all t, we use (3) and equate coefficients of $t^i(1-t)^{4-i}$ to yield a system of equations:

$$\begin{pmatrix} 0 & 0 & 3X_1 & 3Y_1 & -Z_0 \\ 3X_1 & 3Y_1 & 3X_2 & 3Y_2 & -3Z_1 \\ 3X_2 & 3Y_2 & X_3 & Y_3 & -3Z_2 \\ X_3 & Y_3 & 0 & 0 & -Z_3 \end{pmatrix} \begin{pmatrix} P_{0x} \\ P_{0y} \\ P_{1x} \\ P_{1y} \\ q \end{pmatrix} = 0. \tag{9}$$

The system (9) may be solved with Gaussian elimination with each Z_i scaled by a factor L for the purpose of pivot selection. Section 3.2 gives a way to get better accuracy when solving this system.

2.1. The Rotation Strategy

As explained in the introduction, the motivation for rotated coordinates is that there could otherwise be a loss of precision when all control points of the parametric form lie close to a straight line. The barycentric coordinate system suggested by Sederberg [7] may solve some of these problems, but it is difficult to integrate with the implicitization scheme described above and it is not clear how to use it when the curve has points of inflection.

When the control points are close to a straight line, $F(x, y)$ is roughly proportional to the cube of the distance from the line, leading to inaccuracy when the line is not approximately horizontal or vertical. Another way to look at this is that large changes in the ratio of t to $1-t$ correspond to small changes in the ratio of $x(t)$ to $y(t)$, hence P must be almost singular in order for (7) to give the ratio of $1-t$ to $-t$.

In view of the importance of nearly singular P, a geometric interpretation of P helps to choose a system of rotated coordinates. As can be seen from (5), (P_{0x}, P_{0y}) is perpendicular to $\big(x(1) - x_c, y(1) - y_c\big)$ and (P_{1x}, P_{1y}) is perpendicular to $\big(x(0) - x_c, y(0) - y_c\big)$. In fact for all t, vectors

$$(1-t)(-P_{1y}, P_{1x}) + t(-P_{0y}, P_{0x}) \tag{10}$$

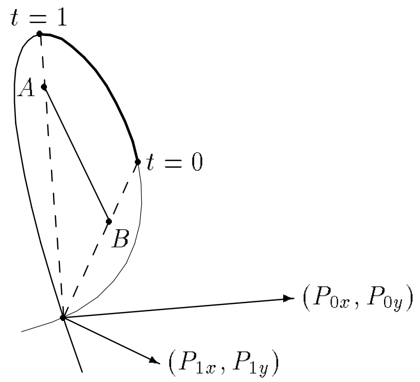


Figure 2: The geometric interpretation of P. When rotated 90°, the vectors (P_{0x}, P_{0y}) and (P_{1x}, P_{1y}) give the displacements from the crossing point to A and B. A point moving at a uniform speed from B to A would always lie on the line from the crossing point to $\big(x(t), y(t)\big)$.

and $\big(x(t) - x_c, y(t) - y_c\big)$ are collinear since multiplying either by $P(t\ \ 1-t)^T$ yields zero. This situation is illustrated in Figure 2.

The error bounds in Section 3 give proof of the appropriateness of our choice of rotated coordinates, but the intuitive basis is that when the choice turns out to be important, the shifted curve $\big(x(t) - x_c, y(t) - y_c\big)$ is approximately a rescaled version of (10). Note that this works whether or not the double point (x_c, y_c) is a crossing point. Thus if the double point and the endpoints of (10) are almost collinear, that direction should be one of the new coordinate axes. It therefore suffices to take the singular value decomposition of P since that tells in what direction a vector is most elongated when multiplied by P.

The singular value decomposition of P may be written

$$\begin{pmatrix} P_{0x} & P_{1x} \\ P_{0y} & P_{1y} \end{pmatrix} = \gamma \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix} \begin{pmatrix} b_1 & b_2 \\ -b_2 & b_1 \end{pmatrix}, \tag{11}$$

where $|\epsilon| \le 1$ and $a_1^2 + a_2^2 = b_1^2 + b_2^2 = 1$. In other words, $P = \gamma AEB$, where γ is a positive scale factor, A and B are rotation matrices, and E is a diagonal matrix. (An efficient way to compute this singular value decomposition appears in Appendix A.)

With the singular value decomposition $P = \gamma AEB$, we use A to define the rotated coordinates (r, s) of a point (x, y):

$$(\ r \quad s\ ) = (\ x \quad y\ )A.$$

Thus $(x\ y)P = \gamma(x\ y)AEB = \gamma(r\ s)EB$, and

$$(\ x \quad y\ )P = \gamma(\ rb_1 - \epsilon s b_2 \quad rb_2 + \epsilon s b_1\ ). \tag{12}$$

2.2. The Implicit Form

We now derive an implicit form for the curve (2) based on rotated coordinates (r, s) as defined by the rotation matrix A in the previous section. The first step is to solve (9) and compute the singular value decomposition (11). To eliminate the parameter, just substitute (12) into (7) and multiply through by $1/\gamma$ to see that the ratio of $1 - t$ to $-t$ is

$$rb_1 - \epsilon s b_2 - \rho : rb_2 + \epsilon s b_1, \tag{13}$$

where

$$\rho = q/\gamma.$$

The implicit form comes from using (13) with a function of t that is zero when $\big(x(t)\ y(t)\big)A = (r\ s)$. One such function is $s\theta(t) - r\phi(t)$, where $\big(\theta(t), \phi(t)\big)$ is the rotated version of $\big(f(t), g(t)\big)$. To facilitate the use of (13), we add another parameter u and define

$$\theta(\tau, u) = 3R_1\tau u^2 + 3R_2\tau^2 u + R_3\tau^3,$$
$$\phi(\tau, u) = 3S_1\tau u^2 + 3S_2\tau^2 u + S_3\tau^3,$$

where

$$(\ R_i \quad S_i\ ) = (\ X_i \quad Y_i\ )A \quad \text{for} \quad i = 1, 2, 3.$$

Thus $\theta(t) = (\tau + u)^{-3}\theta(\tau, u)$ and $\phi(t) = (\tau + u)^{-3}\phi(\tau, u)$ when the ratio $t : 1 - t$ is $\tau : u$. Dividing out the common factor of τ from $\theta(\tau, u)$ and $\phi(\tau, u)$ and using (13) for $-u : \tau$, we obtain the implicit form

$$\frac{s\theta(\tau, u) - r\phi(\tau, u)}{\tau} = 0, \tag{14}$$

where $\tau = rb_2 + \epsilon sb_1$ and $u = -rb_1 + \epsilon sb_2 + \rho$. Expanding as a polynomial in r and s, the implicit form can be written $G(r, s) = 0$, where

$$G(r, s) = \sum_{1 \le m+n \le 3} c_{mn} r^m s^n$$

and

$$\begin{aligned}
c_{30} &= -b_2^2 S_3 + 3b_1 b_2 S_2 - 3b_1^2 S_1,\\
c_{21} &= b_2^2 R_3 - 3b_1 b_2 R_2 + 3b_1^2 R_1 + \epsilon(-2b_1 b_2 S_3 + 3(b_1^2 - b_2^2)S_2 + 6b_1 b_2 S_1),\\
c_{12} &= \epsilon(2b_1 b_2 R_3 + 3(b_2^2 - b_1^2)R_2 - 6b_1 b_2 R_1) - \epsilon^2(b_1^2 S_3 + 3b_1 b_2 S_2 + 3b_2^2 S_1),\\
c_{03} &= \epsilon^2(b_1^2 R_3 + 3b_1 b_2 R_2 + 3b_2^2 R_1),\\
c_{20} &= 3\rho(2b_1 S_1 - b_2 S_2),\\
c_{11} &= 3\rho(-2b_1 R_1 + b_2 R_2 - \epsilon(2b_2 S_1 + b_1 S_2)),\\
c_{02} &= 3\epsilon\rho(2b_2 R_1 + b_1 R_2),\\
c_{10} &= -3\rho^2 S_1,\\
c_{01} &= 3\rho^2 R_1.
\end{aligned} \tag{15}$$

An efficient way to evaluate this is to let

$$c_{30} = -S_{201}, \quad c_{21} = R_{201} + 2\epsilon S_{111}, \quad c_{12} = -2\epsilon R_{111} + \epsilon^2 S_{021}, \quad c_{03} = \epsilon^2 R_{021},$$

$c_{20} = 2\rho S_{101}$, $c_{11} = -2\rho(R_{101} + \epsilon S_{011})$, $c_{02} = 2\rho\epsilon R_{011}$, $c_{10} = -\rho^2 S_{001}$, and $c_{01} = \rho^2 R_{001}$, where

$$\begin{aligned}
R_{001} &= 3R_1, & S_{001} &= 3S_1,\\
R_{002} &= \tfrac{3}{2}R_2, & S_{002} &= \tfrac{3}{2}S_2,\\
R_{101} &= b_1 R_{001} - b_2 R_{002}, & S_{101} &= b_1 S_{001} - b_2 S_{002},\\
R_{102} &= b_1 R_{002} - b_2 R_3, & S_{102} &= b_1 S_{002} - b_2 S_3,\\
R_{011} &= b_2 R_{001} + b_1 R_{002}, & S_{011} &= b_2 S_{001} + b_1 S_{002},\\
R_{012} &= b_2 R_{002} + b_1 R_3, & S_{012} &= b_2 S_{002} + b_1 S_3,\\
R_{201} &= b_1 R_{101} - b_2 R_{102}, & S_{201} &= b_1 S_{101} - b_2 S_{102},\\
R_{111} &= b_2 R_{101} + b_1 R_{102}, & S_{111} &= b_2 S_{101} + b_1 S_{102},\\
R_{021} &= b_2 R_{011} + b_1 R_{012}, & S_{021} &= b_2 S_{011} + b_1 S_{012}.
\end{aligned}$$

2.3. Alternative Cubic Terms

We now have a complete implicitization scheme, but it fails when ρ approaches zero. For example, suppose we are given the parametric form (3) with $Z_0 = Z_1 = Z_2 = Z_3 = 1$, $(X_1, Y_1) = (-1, 1)$, $(X_2, Y_2) = (0, 2)$, and $(X_3, Y_3) = (3, 3)$. Then (9) has the solution $P_{0x} = P_{1x} = P_{1y} = 1$, $P_{0y} = -1$, and $q = 0$. It is easy to check that the singular value decomposition can be written $P = \gamma AEB$, where $\gamma = \sqrt{2}$, A and E are the identity matrix, and

$$B = \begin{pmatrix} b_1 & b_2 \\ -b_2 & b_1 \end{pmatrix}$$

with $b_1 = b_2 = 1/\sqrt{2}$. Since the rotation matrix A is the identity, we have $(R_i, S_i) = (X_i, Y_i)$ for $i = 1, 2, 3$. Using $\epsilon = 1$ and $\rho = q/\gamma$ in (15), we find that all c_{mn} evaluate to zero.

We can avoid the problem by finding alternative expressions for c_{30}, c_{21}, c_{12}, and c_{03} with explicit factors of ρ so that ρ can be factored out of the implicit form $G(r, s) = 0$.

The derivation of an alternative implicit form requires some relationship among the variables of (15). This can be provided by introducing (τ, u) parameters and rotated coordinates into the equation (8) used to derive (9). The rotated version

$$\gamma\begin{pmatrix} \theta(t) & \phi(t) \end{pmatrix} EB \begin{pmatrix} t & 1-t \end{pmatrix}^T - qth(t) = 0$$

is obtained by substituting γAEB for P and $\big(\theta(t)\ \phi(t)\big)$ for $\big(f(t)\ g(t)\big)A$. After multiplying both sides by $1/\gamma$ to give

$$\begin{pmatrix} \theta(t) & \phi(t) \end{pmatrix} EB \begin{pmatrix} t & 1-t \end{pmatrix}^T = \rho th(t),$$

substituting $\tau/(\tau + u)$ for t and multiplying through by $(\tau + u)^4$ yields

$$\begin{pmatrix} \theta(\tau, u) & \phi(\tau, u) \end{pmatrix} EB \begin{pmatrix} \tau & u \end{pmatrix}^T = \rho\tau h(\tau, u), \tag{16}$$

where

$$h(\tau, u) = (\tau + u)^3 h(t) = Z_0 u^3 + 3Z_1 \tau u^2 + 3Z_2 \tau^2 u + Z_3 \tau^3.$$

Note that we can assume we do not have $\epsilon = \rho = 0$ because then (16) forces $\theta(\tau, u)$ to be identically zero so that the desired implicit form is $r^3 = 0$.

In order to use (16) to obtain an alternative version of (14), note that

$$\begin{aligned} \begin{pmatrix} \tau \\ u \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \rho \\ 0 \end{pmatrix} \\ &= QB^T E \begin{pmatrix} r \\ s \end{pmatrix} - Q \begin{pmatrix} \rho \\ 0 \end{pmatrix}, \end{aligned}$$

where Q is a rotation matrix, and τ and u are as chosen in (14). Substituting this into (16) yields

$$\begin{pmatrix} \theta(\tau, u) & \phi(\tau, u) \end{pmatrix} EBQB^T E \begin{pmatrix} r \\ s \end{pmatrix} = \rho\tau h(\tau, u) + \begin{pmatrix} \theta(\tau, u) & \phi(\tau, u) \end{pmatrix} EBQ \begin{pmatrix} \rho \\ 0 \end{pmatrix}.$$

Since $EBQB^T E = EQBB^T E = EQE = \epsilon Q$, we have

$$\begin{aligned} \epsilon\begin{pmatrix} \theta(\tau, u) & \phi(\tau, u) \end{pmatrix} Q \begin{pmatrix} r & s \end{pmatrix}^T &= \rho\tau h(\tau, u) + \begin{pmatrix} \theta(\tau, u) & \phi(\tau, u) \end{pmatrix} EBQ \begin{pmatrix} \rho & 0 \end{pmatrix}^T \\ &= \rho\tau h(\tau, u) - \rho\big(b_2\theta(\tau, u) + \epsilon b_1 \phi(\tau, u)\big). \end{aligned}$$

Multiplying out the left-hand side and dividing both sides by $\epsilon\tau$ yields

$$\frac{s\theta(\tau, u) - r\phi(\tau, u)}{\tau} = \frac{\rho}{\epsilon}\left(h(\tau, u) - \frac{b_2\theta(\tau, u) + \epsilon b_1\phi(\tau, u)}{\tau}\right), \tag{17}$$

where $\tau = rb_2 + \epsilon s b_1$ and $u = -rb_1 + \epsilon s b_2 + \rho$ as in (14).

The alternative implicit form obtained by substituting (17) into (14) has the desired explicit factor of ρ. By expanding the right-hand side of (17) as a polynomial in r and s, we obtain expressions for the coefficients like (15), but with ρ factored out. The coefficients of linear and quadratic terms turn out to be rather complicated and of little interest to us, but the cubic terms are simpler because they all come from $\frac{\rho}{\epsilon}h(\tau, u)$; i.e.,

$$\frac{\rho}{\epsilon}h(rb_2 + \epsilon s b_1,\ -rb_1 + \epsilon s b_2) = c_{30}r^3 + c_{21}r^2 s + c_{12} r s^2 + c_{03} s^3,$$

where

$$\begin{aligned}
c_{30} &= \rho\epsilon^{-1}(-b_1^3 Z_0 + 3b_1^2 b_2 Z_1 - 3b_1 b_2^2 Z_2 + b_2^3 Z_3) \\
c_{21} &= 3\rho(b_1^2 b_2 Z_0 + b_1(b_1^2 - 2b_2^2)Z_1 + b_2(b_2^2 - 2b_1^2)Z_2 + b_1 b_2^3 Z_3) \\
c_{12} &= 3\rho\epsilon(-b_1 b_2^2 Z_0 + b_2(b_2^2 - 2b_1^2)Z_1 - b_1(b_1^2 - 2b_2^2)Z_2 + b_1^2 b_2 Z_3) \\
c_{03} &= \rho\epsilon^2(b_2^3 Z_0 + 3b_1 b_2^2 Z_1 + 3b_1^2 b_2 Z_2 + b_1^3 Z_3).
\end{aligned} \tag{18}$$

An efficient way to evaluate this is to let $c_{30} = \rho\epsilon^{-1}B_{300}$, $c_{21} = 3\rho B_{210}$, $c_{12} = 3\rho\epsilon B_{120}$, and $c_{03} = \rho\epsilon^2 B_{030}$, where

$$\begin{aligned}
B_{100} &= -b_1 Z_0 + b_2 Z_1, & B_{010} &= b_2 Z_0 + b_1 Z_1, \\
B_{101} &= -b_1 Z_1 + b_2 Z_2, & B_{011} &= b_2 Z_1 + b_1 Z_2, \\
B_{102} &= -b_1 Z_2 + b_2 Z_3, & B_{012} &= b_2 Z_2 + b_1 Z_3, \\
B_{200} &= -b_1 B_{100} + b_2 B_{101}, & B_{020} &= b_2 B_{010} + b_1 B_{011}, \\
B_{201} &= -b_1 B_{101} + b_2 B_{102}, & B_{021} &= b_2 B_{011} + b_1 B_{012}, \\
B_{300} &= -b_1 B_{200} + b_2 B_{201}, & B_{210} &= b_2 B_{200} + b_1 B_{201}, \\
B_{120} &= -b_1 B_{020} + b_2 B_{021}, & B_{030} &= b_2 B_{020} + b_1 B_{021}.
\end{aligned}$$

In the implicit form $G(r, s) = 0$, the polynomial $G(r, s)$ is a sum of terms of the form $c_{mn} r^m s^n$ for $1 \le m + n \le 3$, where any desired combination of the expressions (15) and (18) may be used to evaluate the coefficients.

When ρ is small and we need to evaluate the coefficients c_{mn}/ρ in the implicit form $G(r, s)/\rho$, we avoid factors of ρ in the denominator by using (18) instead of (15) for c_{30}/ρ, c_{21}/ρ, c_{12}/ρ, and c_{03}/ρ. As explained in Section 3, the phrase "ρ is small" can be expressed as a specific comparison involving the magnitudes of ρ and ϵ. When the comparison succeeds, we use the alternative cubic terms along with the material in the next section.

2.4. The Double Point Centered Form

It is well known that if (r_c, s_c) is the double point of a cubic curve with implicit form $G(r, s) = 0$, then $G(r_c, s_c) = 0$ and (r_c, s_c) is a saddle point if there is a crossing there, or a relative maximum or minimum if there is no crossing. Either way $G(r, s)$ is very flat in the neighborhood of the double point, and as Figure 3 shows, this magnifies roundoff errors that arise when the coefficients of $G(r, s)$ are represented imprecisely. This section shows how to avoid such inaccuracies when the double point is on or near the curve segment specified by the given Bézier control points.

We can avoid inaccuracy near the double point (r_c, s_c) by finding r_c and s_c and giving $G(r, s)$ as a polynomial in $r - r_c$ and $s - s_c$. Hence everything depends on evaluating r_c and s_c and the coefficients of this polynomial so that the curve defined by the implicit form is as accurate as possible. How can this be done when the computation of (r_c, s_c) is so ill-conditioned? The key is that (6) shows that the distance from the origin to the double point is directly related to q, hence the computed value of q largely determines the position of the double point in difficult cases such as

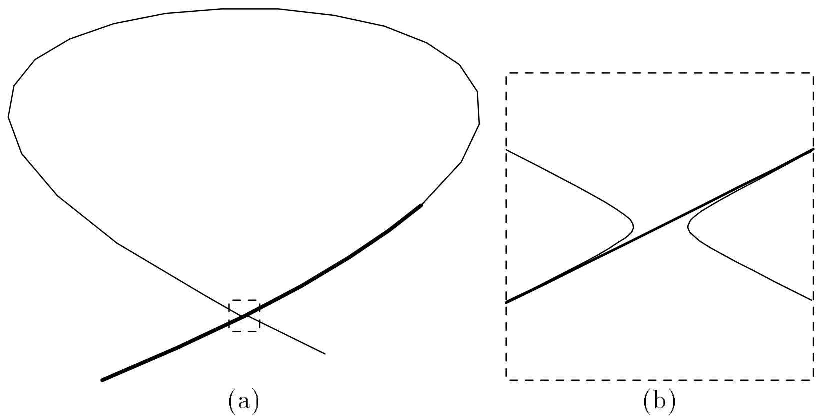


Figure 3: (a) The polynomial cubic where the corresponding space curve has Bézier control points $(0,0,1),(263,110,1),(427,205,1)$, and $(519,285,1)$ (heavy line) and zeros of an approximate implicit form with coefficients rounded to four decimal digits (thin line); (b) A closeup of the portion of part (a) within the dashed square where the error is particularly large.

shown in Figure 1. In other words, q is difficult to compute accurately, but the implicit curve can still be accurate as long as subsequent calculations all use the same value for q.

Letting $P = \gamma AEB$ and $(r_c, s_c) = (x_c, y_c)A$ in (6) yields

$$\gamma \begin{pmatrix} r_c & s_c \end{pmatrix} EB \begin{pmatrix} t & 1-t \end{pmatrix}^T = qt.$$

Since this must hold for all t, we have

$$\gamma \begin{pmatrix} r_c & s_c \end{pmatrix} EB = \begin{pmatrix} q & 0 \end{pmatrix}$$

or

$$\begin{pmatrix} r_c & s_c \end{pmatrix} = \begin{pmatrix} \rho & 0 \end{pmatrix} B^T E^{-1} = \begin{pmatrix} \rho b_1, -\rho b_2/\epsilon \end{pmatrix}. \tag{19}$$

Since it is known that G and its first derivatives are zero at (r_c, s_c), we have

$$G(r,s) = \sum_{1 \le m+n \le 3} c_{mn} r^m s^n = \sum_{2 \le m+n \le 3} c'_{mn} (r - s_c)^m (s - s_c)^n,$$

and therefore

$$c'_{30} = c_{30}, \quad c'_{21} = c_{21}, \quad c'_{12} = c_{12}, \quad c'_{03} = c_{03} \tag{20}$$

and

$$\begin{aligned} c'_{20} &= c_{20} + 3r_c c_{30} + s_c c_{21}, \\ c'_{11} &= c_{11} + 2r_c c_{21} + 2s_c c_{12}, \\ c'_{02} &= c_{02} + r_c c_{12} + 3s_c c_{03}. \end{aligned}$$

Substituting the for r_c and s_c as directed by (19) yields

$$\begin{aligned} c'_{20} &= c_{20} + 3\rho b_1 c_{30} - \rho b_2 c_{21}/\epsilon, \\ c'_{11} &= c_{11} + 2\rho b_1 c_{21} - 2\rho b_2 c_{12}/\epsilon, \\ c'_{02} &= c_{02} + \rho b_1 c_{12} - 3\rho b_2 c_{03}/\epsilon. \end{aligned}$$

Using (18) and the fact that $b_1^2 + b_2^2 = 1$, this becomes

$$\begin{aligned} c'_{20} &= c_{20} - 3\rho^2 \epsilon^{-1} (b_1^2 Z_0 - 2b_1 b_2 Z_1 + b_2^2 Z_2), \\ c'_{11} &= c_{11} - 6\rho^2 (-b_1 b_2 Z_0 + (b_2^2 - b_1^2) Z_1 + b_1 b_2 Z_2), \\ c'_{02} &= c_{02} - 3\rho^2 \epsilon (b_2^2 Z_0 + 2b_1 b_2 Z_1 + b_1^2 Z_2). \end{aligned} \tag{21}$$

The complete double point centered form has coefficients (20) and (21), where c_{20}, c_{11}, and c_{02} are as given by (15) and the other c_{mn} are as given by (18). It is particularly convenient that the efficient evaluation scheme for (18) given in Section 2.3 allows (21) to be evaluated as

$$\begin{aligned} c'_{20} &= c_{20} - 3\rho\epsilon^{-1}B_{200}, \\ c'_{20} &= c_{11} - 3\rho B_{110}, \\ c'_{20} &= c_{02} - 3\rho\epsilon B_{020}, \end{aligned}$$

where B_{200} and B_{020} are as in Section 2.3, and $B_{110} = b_2 B_{100} + b_1 B_{101}$.

3. Numerical Stability

In this section, we use backward error analysis to compare the given parametric curve with the implicitly-defined curve obtained by evaluating expressions from Section 2 with floating point arithmetic. Basic arithmetic operations are assumed to produce results with relative error at most δ, where δ is the machine precision.

A two stage approach helps to deal with inaccuracy in locating the double point and determining the four parameters that appear in (19). We analyze the initial stage of computation by finding a perturbed version of the parametric form that corresponds to the computed values of ϵ, ρ, b_1, and b_2. Thus in the computation of these four parameters and the R_i, S_i and Z_i that determine $\theta(t)$, $\phi(t)$, and $h(t)$, we obtain an approximate solution to a perturbed problem. The total error in the curve represented by the computed implicit form is the size of the perturbation plus whatever errors arise from doing the rest of the computation with the computed control points $\hat{R}_i$, $\hat{S}_i$ and the actual Z_i instead of with the perturbed versions of the control points. In other words the perturbation given in Section 3.1 applies to the input to final stage of computation.

We want the magnitude of the perturbation in Z_i to be small compared to

$$h_{\min} = \min_{0 \le t \le 1} |h(t)|$$

and the magnitude of perturbations in X_i and Y_i to be small compared to

$$B_{xy} = (1 + 10\delta) \max_{i=1,2,3} \sqrt{X_i^2 + Y_i^2}.$$

(The factor of $1 + 10\delta$ ensures that $\sqrt{\hat{R}_i^2 + \hat{S}_i^2} < B_{xy}$). The effect of these perturbations on the curve itself should be small compared to

$$L = B_{xy}/h_{\min}$$

since this is a bound on the overall size of the curve as measured by the maximum distance from the starting point.

Since we also need a bound on the maximum magnitude of the control points Z_i, we introduce the constant

$$\lambda = \frac{\max_{i=0,1,2,3} |Z_i|}{h_{\min}}.$$

The constant λ is a measure of how "well-behaved" $h(t)$ is. The best case is that of a polynomial cubic curve where $h(t)$ is constant and $\lambda = 1$. In an extreme case of bad behavior such as that shown in Figure 4, $h_{\min}$ is relatively small and errors are greatly magnified when $h(t) \approx h_{\min}$. The use of λ in the error bounds accounts for the fact that the curve can be highly sensitive to small relative changes in the input when λ is large. By using λ we make the hidden assumption that $h(t)$ does not cross zero, hence we might as well assume that $h(t)$ is always positive when $0 \le t \le 1$.

The purpose of Sections 3.1–3.4 is to prove the following numerical stability theorem. It is expressed as an asymptotic result as $\delta \to 0$, but the constant in the O is really just a number that could be computed by doing the analysis more carefully.

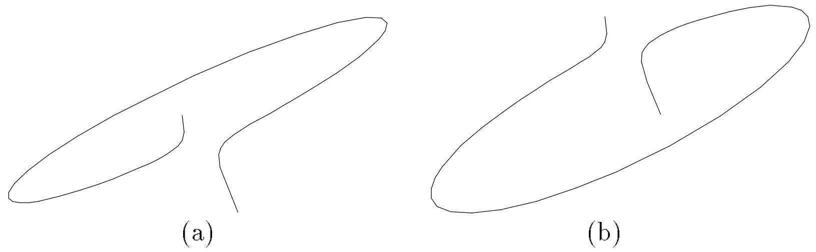


Figure 4: (a) The curve generated when $(X_1, X_2, X_3) = (14.04, -56.16, 157.4)$, $(Y_1, Y_2, Y_3) = (-45.88, 129.84, -274.76)$ and $(Z_0, Z_1, Z_2, Z_3) = (37.21, -28.67, 3.05, 54.29)$; (b) the effect of changing Y_3 from -274.76 to -275.37. The effect is magnified because $h(t)$ achieves a minimum value of about 0.004.

Theorem 3.1 *Let the implicit form of a rational cubic curve be computed with machine precision δ as described in Section 2 using the ratio of the computed parameters $\hat{\rho} : \hat{\epsilon}$ to decide whether to use the alternative forms of Sections 2.3 and 2.4. In the coordinate system of the computed rotation matrix $\left(\begin{smallmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{smallmatrix}\right)$, any point (r, s) on the original curve has a neighbor $(r+\Delta r, s+\Delta s)$ that satisfies the computed implicit form, where*

$$\Delta r = O\left(\lambda^3 \left(\lambda^2 + \frac{\lambda\delta}{\lambda\,|\hat{\epsilon}| + |\hat{\rho}|\,/L}\right) L\delta\right) \quad \text{and} \quad \Delta s = O\left(\lambda^3 \left(\lambda^2 + \frac{1}{|\lambda\hat{\epsilon}| + |\hat{\rho}|\,/L}\right) L\delta\right)$$

for sufficiently small δ. With the iterative improvement scheme of Section 3.2, the displacement can be reduced so that

$$\sqrt{(\Delta r)^2 + (\Delta s)^2} = O\left(\lambda^3 \left(\lambda^2 + \frac{\delta}{\lambda\,|\hat{\epsilon}| + |\hat{\rho}|\,/L}\right) L\delta\right).$$

The theorem bounds the errors in the implicit curve in terms of the scale parameter L and the machine precision δ. The high powers of λ appear to be pessimistic in practice, but some dependence on λ is definitely necessary to deal with cases such as Figure 4. The only troubling terms are the denominators that depend on $\hat{\epsilon}$ and $\hat{\rho}/L$. These are best understood with the aid of Theorem B.2 from Appendix B. Combining this with Lemma 3.6 from Section 3.1

$$\max_{i=1,2,3} \left|\hat{R}_i\right| = O((|\hat{\epsilon}| + \lambda\,|\hat{\rho}|\,/L + \lambda\delta)B_{xy}). \tag{22}$$

The meaning of (22) is that when the denominators in Theorem 3.1 are much less than one, the range of r coordinates is correspondingly smaller than the range of s coordinates. Thus the denominators cannot be as small as $\lambda\delta$ except when $r^3 = 0$ is a sufficiently accurate implicit form.

The

$$O\left(\frac{\lambda^3 L\delta}{|\lambda\hat{\epsilon}| + |\hat{\rho}|\,/L}\right)$$

term in the equation for Δs still presents a problem, but this is the reason for the iterative improvement scheme of Section 3.2. Note that this is not necessary when the curve slope $\frac{dr}{ds}$ reflects the overall dimensions implied by (22), since the component of $(\Delta r, \Delta s)$ perpendicular to the curve is then on the order of

$$|\Delta r| + (|\hat{\epsilon}| + \lambda\,|\hat{\rho}|\,/L + \lambda\delta)\,|\Delta s|\,.$$

Theorem 3.1 is a simple consequence of the four lemmas stated below. Lemmas 3.2 and 3.3 give bounds on the perturbations necessary to account for the errors in $\hat{\epsilon}$, $\hat{\rho}$, $\hat{b}_1$, and $\hat{b}_2$. The proofs are given in Sections 3.1 and 3.2 but the casual reader will probably want to skip them and go on to Section 4.

Lemma 3.2 *The perturbation of the computed control points $\hat{R}_i$, $\hat{S}_i$, Z_i necessary to make them agree with the computed parameters $\hat{\epsilon}$, $\hat{\rho}$, $\hat{b}_1$, and $\hat{b}_2$ effects $r = \theta(t)/h(t)$ and $s = \phi(t)/h(t)$ by*

$$\Delta r = O(\lambda^2 \delta L) \quad \text{and} \quad \Delta s \le O\left(\frac{\lambda \delta L}{|\hat{\epsilon}| + |\hat{\rho}|/L}\right).$$

Lemma 3.3 *With the iterative improvement scheme of Section 3.2, the perturbation of $\hat{R}_i$, $\hat{S}_i$, and Z_i necessary to make them agree with $\hat{\epsilon}$, $\hat{\rho}$, $\hat{b}_1$, and $\hat{b}_2$ effects $r = \theta(t)/h(t)$ and $s = \phi(t)/h(t)$ by*

$$\Delta r = O((\lambda |\hat{\epsilon}| + \lambda^2 |\hat{\rho}|/L + \lambda^2 \delta)\delta L) \quad \text{and} \quad \Delta s = O(\lambda \delta L).$$

The final stage of the implicitization process involves comparing $|\hat{\rho}|$ to $|\hat{\epsilon}|$ to decide whether to use the alternative cubic terms and the double point centered form. To this end, it is useful to have a bound on $|\rho|$ analogous to the bound $|\epsilon| \le 1$ given in Section 2.1.

According to (9), $q = (X_3 P_{0x} + Y_3 P_{0y})/Z_3$ hence

$$|q| \le L\sqrt{P_{0x}^2 + P_{0y}^2} \quad \text{and} \quad |\rho| \le L\gamma^{-1}\sqrt{P_{0x}^2 + P_{0y}^2}.$$

It follows from (11) that $\sqrt{P_{0x}^2 + P_{0y}^2} \le \gamma$ and thus

$$|\rho| \le L.$$

When $|\hat{\rho}|$ is sufficiently greater than $L|\hat{\epsilon}|$, Lemma 3.4 below gives a good error bound for implicitization without the alternative cubic terms or the double point centered form. The proof in Section 3.3 requires δ to be "sufficiently small" because of the way it uses an upper bound on the error in $G(r,s)$ to estimate the effect on the curve $G(r,s) = 0$. Section 3.3 explains that it probably suffices to have

$$\delta \le c/\lambda^k$$

for some constant c and some small integer k.

Lemma 3.4 *When $|\hat{\rho}| \ge 33(\lambda - \frac{1}{9})L|\hat{\epsilon}|$ and δ is sufficiently small, all points on the perturbed curve of Lemma 3.2 are within*

$$O(\lambda^5 L \delta)$$

of a zero of the implicit form computed as described in Section 2.2.

When ρ is larger than the bound in Lemma 3.4, the final stage of computation is to determine the double-point centered form as described in Section 2.4. Lemma 3.5 gives error bounds in this case. The proof in Section 3.4 involves finding a rational parameterization of the computed implicit form and comparing this to the perturbed curves of Lemmas 3.2 and 3.3. Once this is done, Theorem 3.1 follows by combining the error bounds in Lemmas 3.2–3.5.

Lemma 3.5 *When $|\hat{\rho}| \le 33(\lambda - \frac{1}{9})L|\hat{\epsilon}|$, there is a one to one correspondence between points on the implicit form computed according to Sections 2.3 and 2.4 and points on the perturbed curve of Lemma 3.2 where corresponding points differ by*

$$|\Delta r| = O\left(\lambda^2\left(1 + \frac{\lambda\delta}{\hat{\epsilon}}\right)L\delta\right) \quad \text{and} \quad |\Delta s| = O\left(\frac{\lambda^2 L\delta}{\hat{\epsilon}}\right).$$

With the iterative improvement scheme of Section 3.2, the difference is reduced to

$$|\Delta r| = O\left(\lambda^3\left(\hat{\epsilon} + \lambda\delta + \frac{\delta^2}{\hat{\epsilon}}\right)L\delta\right) \quad \text{and} \quad |\Delta s| = O\left(\lambda^2\left(\lambda + \frac{\delta}{\hat{\epsilon}}\right)L\delta\right).$$

3.1. Analysis of the Initial Stage

The purpose of this section is to prove Lemma 3.2 as explained at the beginning of Section 3. To do this, we must decide how closely the curve represented by the rotated Bézier control points (R_i, S_i, Z_i) corresponds to the parameters obtained by solving the system (9) and finding the singular value decomposition (11). This is mostly a matter of getting bounds on the residual in (9) and (23) and then finding perturbations of the Z_i, R_i, and S_i that eliminate the residual in (23).

Lemma 3.6 *It is possible to solve (9), find the singular value decomposition (11), and evaluate $\rho = q/\gamma$ in floating point of precision δ so that the infinity norm of the residual*

$$\mathcal{R} = \begin{pmatrix} 0 & 0 & 3R_1 & 3S_1 & -LZ_0 \\ 3R_1 & 3S_1 & 3R_2 & 3S_2 & -3LZ_1 \\ 3R_2 & 3S_2 & R_3 & S_3 & -3LZ_2 \\ R_3 & S_3 & 0 & 0 & -LZ_3 \end{pmatrix} \begin{pmatrix} b_1 \\ -b_2\epsilon \\ b_2 \\ b_1\epsilon \\ \rho/L \end{pmatrix} \tag{23}$$

is $O(\lambda B_{xy}\delta)$.

Proof. Begin by solving (9) with the Z_i column scaled by L; i.e., use Gaussian elimination with complete pivoting to find an approximate solution vector $\tilde{V}$ to the system

$$\begin{pmatrix} 0 & 0 & 3X_1 & 3Y_1 & -LZ_0 \\ 3X_1 & 3Y_1 & 3X_2 & 3Y_2 & -3LZ_1 \\ 3X_2 & 3Y_2 & X_3 & Y_3 & -3LZ_2 \\ X_3 & Y_3 & 0 & 0 & -LZ_3 \end{pmatrix} \begin{pmatrix} P_{0x} \\ P_{0y} \\ P_{1x} \\ P_{1y} \\ q/L \end{pmatrix} = 0,$$

where the residual has infinity norm $O\big(\lambda B_{xy}\|\tilde{V}\|_\infty\delta\big)$.

The next step is to take the singular value decomposition of the matrix $\tilde{P}$ whose entries are the first four entries of $\tilde{V}$. The computed decomposition $\hat{\gamma}\hat{A}\hat{E}\hat{B}$ has $\|\tilde{P} - \hat{\gamma}\hat{A}\hat{E}\hat{B}\|_\infty = O(\delta\|\tilde{P}\|_\infty)$, and substituting $\hat{\gamma}\hat{A}\hat{E}\hat{B}$ for $\tilde{P}$ in $\tilde{V}$ adds $O\big(B_{xy}\|\tilde{P}\|_\infty\delta\big)$ to the infinity norm of the residual. This estimated residual can be written

$$\hat{\gamma}\begin{pmatrix} 0 & 0 & 3\tilde{R}_1 & 3\tilde{S}_1 & -LZ_0 \\ 3\tilde{R}_1 & 3\tilde{S}_1 & 3\tilde{R}_2 & 3\tilde{S}_2 & -3LZ_1 \\ 3\tilde{R}_2 & 3\tilde{S}_2 & \tilde{R}_3 & \tilde{S}_3 & -3LZ_2 \\ \tilde{R}_3 & \tilde{S}_3 & 0 & 0 & -LZ_3 \end{pmatrix} \begin{pmatrix} \hat{b}_1 \\ -\hat{b}_2\hat{\epsilon} \\ \hat{b}_2 \\ \hat{b}_1\hat{\epsilon} \\ \tilde{q}/(\hat{\gamma}L) \end{pmatrix}, \tag{24}$$

where $\tilde{R}_i = \hat{a}_1X_i + \hat{a}_2Y_i$, $\tilde{S}_i = \hat{a}_1Y_i - \hat{a}_2X_i$, $\hat{a}_1$ and $\hat{a}_2$ are the entries of $\hat{A}$, $\hat{b}_1$ and $\hat{b}_2$ are the entries of $\hat{B}$, and $\hat{\epsilon}$ and $\tilde{q}/L$ are entries of $\hat{E}$ and $\tilde{V}$ respectively.

The estimated residual is identical to the computed version of the left-hand side of (23) except for the factor of $\hat{\gamma}$, the use of $\tilde{q}/\hat{\gamma}$ instead of the computed quotient $\hat{\rho}$, and the use of $\tilde{R}_i$ and $\tilde{S}_i$ in place of the computed values $\hat{R}_i$ and $\hat{S}_i$. Since switching to $\hat{\rho}$ adds a term with infinity norm $O(\lambda B_{xy}\delta)$ and switching to $\hat{R}_i$ and $\hat{S}_i$ adds a term with norm $O(B_{xy}\delta)$, the residual

$$\mathcal{R} = \begin{pmatrix} 0 & 0 & 3\hat{R}_1 & 3\hat{S}_1 & -LZ_0 \\ 3\hat{R}_1 & 3\hat{S}_1 & 3\hat{R}_2 & 3\hat{S}_2 & -3LZ_1 \\ 3\hat{R}_2 & 3\hat{S}_2 & \hat{R}_3 & \hat{S}_3 & -3LZ_2 \\ \hat{R}_3 & \hat{S}_3 & 0 & 0 & -LZ_3 \end{pmatrix} \begin{pmatrix} \hat{b}_1 \\ -\hat{b}_2\hat{\epsilon} \\ \hat{b}_2 \\ \hat{b}_1\hat{\epsilon} \\ \hat{\rho}/L \end{pmatrix} \tag{25}$$

has infinity norm

$$\|\mathcal{R}\|_\infty = O\left(\frac{B_{xy}\|\tilde{P}\|_\infty\delta + \lambda B_{xy}\|\tilde{V}\|_\infty\delta}{\hat{\gamma}} + \lambda B_{xy}\delta\right) = O(\lambda B_{xy}\delta). \tag{26}$$

□

We now need to decide what perturbations of the $\hat{R}_i$, $\hat{S}_i$, and Z_i parameters are necessary to cancel this residual. If $\hat{\rho}/L$ has absolute value greater than $|\hat{\epsilon}|$, it suffices to let

$$\hat{\rho}(\; \Delta Z_0 \quad 3\Delta Z_1 \quad 3\Delta Z_2 \quad \Delta Z_3 \;)^T = -\mathfrak{R}.$$

This produces a perturbation of the $\hat{Z}_i$ with an infinity norm at most $|\hat{\rho}|^{-1} \|\mathfrak{R}\|_\infty$.

When $|\hat{\epsilon}| > |\hat{\rho}|/L$, it is better to perturb the $\hat{R}_i$ and $\hat{S}_i$. The terms of $\mathfrak{R}$ that depend on the $\hat{R}_i$ and $\hat{S}_i$ can be written

$$Q(\; \hat{R}_1 \quad \hat{R}_2 \quad \hat{R}_3 \quad \epsilon\hat{S}_1 \quad \epsilon\hat{S}_2 \quad \epsilon\hat{S}_3 \;)^T,$$

where

$$Q = \begin{pmatrix} 3\hat{b}_2 & 0 & 0 & 3\hat{b}_2 & 0 & 0 \\ 3\hat{b}_1 & 3\hat{b}_2 & 0 & -3\hat{b}_1 & 3\hat{b}_2 & 0 \\ 0 & 3\hat{b}_1 & \hat{b}_2 & 0 & -3\hat{b}_1 & \hat{b}_2 \\ 0 & 0 & \hat{b}_1 & 0 & 0 & -\hat{b}_1 \end{pmatrix}.$$

Thus the perturbation vector should have the property that multiplying it on the left by Q yields $-\mathfrak{R}$. We therefore let

$$(\; \Delta R_1 \quad \Delta R_2 \quad \Delta R_3 \quad \hat{\epsilon}\Delta S_1 \quad \hat{\epsilon}\Delta S_2 \quad \hat{\epsilon}\Delta S_3 \;) = Q^T(QQ^T)^{-1}\mathfrak{R}.$$

Since $\hat{b}_1^2 + \hat{b}_2^2 = 1 + O(\delta)$ for machine precision δ, the product QQ^T is a diagonal matrix with nonzero entries approximately 9, 18, 10, and 1. Thus when δ is reasonable, $\left\|Q^T(QQ^T)^{-1}\right\|_\infty < \frac{11}{10}$ and the perturbations of $\hat{R}_i$ have infinity norm at most $\frac{11}{10}\|\mathfrak{R}\|$, while those for $\hat{S}_i$ have norm $\le \frac{11}{10}|\hat{\epsilon}|^{-1}\|\mathfrak{R}\|$. Thus we have proved the following lemma:

Lemma 3.7 *If the residual in (23) is $\mathfrak{R}$, then (23) is satisfied by a perturbation where*

$$|\Delta Z_i| \le \frac{\|\mathfrak{R}\|_\infty}{|\hat{\rho}|} \quad \text{and} \quad \Delta R_i = \Delta S_i = 0, \tag{27}$$

if $|\hat{\epsilon}| \le |\hat{\rho}|/L$, and

$$|\Delta R_i| \le \frac{11}{10}\|\mathfrak{R}\|_\infty, \quad |\Delta S_i| \le \frac{11\|\mathfrak{R}\|_\infty}{10|\hat{\epsilon}|}, \quad \text{and} \quad \Delta Z_i = 0. \tag{28}$$

otherwise.

The effect of the perturbation (27) on r and s at some time t_0 is to change $h(t_0)$ by an amount no more than $\|\mathfrak{R}\|_\infty / |\hat{\rho}|$. Dividing by $h_{\min}$ and noting that $L = B_{xy}/h_{\min}$, the relative change in $h(t_0)$ is at most $L\|\mathfrak{R}\|_\infty/(|\hat{\rho}|B_{xy})$. Using $|s| \le B_{xy}/h_{\min}$ and $|r| \le \hat{B}_r/h_{\min}$ with the bound on $\hat{B}_r$ from Theorem B.2,

$$|\Delta r| \le \left(7.95\lambda + 11.25\frac{|\hat{\epsilon}|L}{|\hat{\rho}|} + 2.65\frac{L\|\mathfrak{R}\|_\infty}{|\hat{\rho}|B_{xy}}\right)\frac{\|\mathfrak{R}\|_\infty}{h_{\min}} \quad \text{and} \quad |\Delta s| \le \frac{L\|\mathfrak{R}\|_\infty}{|\hat{\rho}|h_{\min}}.$$

When $|\hat{\epsilon}| > |\hat{\rho}|/L$, the effects of (28) on $r = \theta(t_0)/h(t_0)$ and $s = \phi(t_0)/h(t_0)$ are

$$|\Delta r| \le \frac{11\|\mathfrak{R}\|_\infty}{10h_{\min}} \quad \text{and} \quad |\Delta s| \le \frac{11\|\mathfrak{R}\|_\infty}{10|\hat{\epsilon}|h_{\min}}.$$

Taking the maximum possible value of the bounds for all values of the $|\hat{\epsilon}|$ to $|\hat{\rho}|$ ratio yields overall bounds of

$$\begin{aligned} |\Delta r| &\le \left(7.95\lambda + 11.25 + \frac{2.65\|\mathfrak{R}\|_\infty}{B_{xy}\max(|\hat{\epsilon}|, |\hat{\rho}|/L)}\right)\frac{\|\mathfrak{R}\|_\infty}{h_{\min}} \\ |\Delta s| &\le \frac{\|\mathfrak{R}\|_\infty}{\max(\frac{10}{11}|\hat{\epsilon}|, |\hat{\rho}|/L)h_{\min}} \end{aligned} \tag{29}$$

on the size of the perturbations necessary to make parametric curve agree with the computed values $\hat{b}_1$, $\hat{b}_2$, $\hat{\epsilon}$, and $\hat{\rho}$. Combining this with Lemma 3.6 completes the proof of Lemma 3.2.

3.2. The Initial Stage with Iterative Improvement

It would be nice to have a stronger version of Lemma 3.2 that guarantees a small perturbation even when the slope $\frac{ds}{dr}$ is much less than B_{xy}/B_r where $B_r = \max(R_1, R_2, R_3)$ as in Appendix B. This requires an iterative improvement scheme is needed to control the error in the neighborhood of "flat spots" in curves such as those in Figure 5 where $B_r \ll B_{xy}$ and the derivative $s'(t)$ passes through zero.

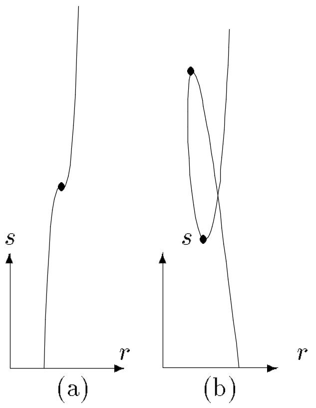


Figure 5: Curves where $B_r \ll B_{xy}$ yet there are points (marked by dots) where the $\frac{ds}{dr}$ slope is zero.

One way to improve the situation for curves such as those in Figure 5 where (29) sometimes allows relatively large errors is to compute better values for $\hat{b}_1$, $\hat{b}_2$, $\hat{\epsilon}$, and $\hat{\rho}$. The basic idea is that the components of the residual (25) can be evaluated to within

$$\begin{aligned} O((\hat{B}_r + \epsilon\hat{B}_{xy} + \lambda h_{\min}\hat{\rho})\delta) &= O(\hat{B}_r\delta + (\hat{\epsilon} + \lambda\hat{\rho}/L)B_{xy}\delta) \\ &= O(\delta\,\|\mathcal{R}\|_\infty + (\hat{\epsilon} + \lambda\hat{\rho}/L)B_{xy}\delta) \\ &= O((\hat{\epsilon} + \lambda\hat{\rho}/L + \lambda\delta)B_{xy}\delta) \qquad (30) \end{aligned}$$

so there is some hope of adjusting the parameters to reduce the residual to this magnitude.

When $\hat{B}_r = 0$ it suffices to let $\rho = \epsilon = 0$; otherwise we do the adjustment by essentially repeating the original calculation with $\hat{R}_i$ and $\hat{S}_i$ in place of X_i and Y_i for $i = 1, 2, 3$. The only modification is that we scale the new X_i by the old $B_{xy}/\hat{B}_r$ when solving the linear system (9).

In other words, an iterative improvement step begins by using Gaussian elimination with complete pivoting to find an approximate solution vector $\tilde{V}$ to the system

$$\begin{pmatrix} 0 & 0 & 3\alpha X_1 & 3Y_1 & -LZ_0/\lambda \\ 3\alpha X_1 & 3Y_1 & 3\alpha X_2 & 3Y_2 & -3LZ_1/\lambda \\ 3\alpha X_2 & 3Y_2 & \alpha X_3 & Y_3 & -3LZ_2/\lambda \\ \alpha X_3 & Y_3 & 0 & 0 & -LZ_3/\lambda \end{pmatrix} \begin{pmatrix} P_{0x}/\alpha \\ P_{0y} \\ P_{1x}/\alpha \\ P_{1y} \\ \lambda q/L \end{pmatrix} = 0, \qquad (31)$$

where $\alpha = B_{xy}/B_x$ and $B_x = \max_{i=1,2,3}|X_i|$ is the old $\hat{B}_r$. After multiplying the appropriate elements of $\tilde{V}$ by α and L/λ to find P_{0x}, P_{1x}, and q, the new implicit form can be computed exactly as in Sections 2.1–2.4. The following lemma gives error bounds for this process:

Lemma 3.8 *Let $\hat{\epsilon}_0$ and $\hat{\rho}_0$ be the computed versions of ϵ and ρ before an iterative improvement step, and $\hat{\epsilon}$ and $\hat{\rho}$ be the corresponding parameters afterward. After the iterative improvement step, the residual $\mathcal{R}$ from (23) satisfies*

$$\|\mathcal{R}\|_\infty = O\left(\left(\hat{\epsilon} + \frac{\lambda\hat{\rho}}{L} + \hat{\epsilon}_0 + \frac{\lambda\hat{\rho}_0}{L} + \lambda\delta\right) B_{xy}\delta\right).$$

Proof. The residual in (31) after Gaussian elimination with complete pivoting has infinity norm that is $O(\|\tilde{V}\|_\infty B_{xy}\delta)$.

Let D be the diagonal matrix diag$(1, 1/\alpha)$ so that for P as in (11), the entries of DP are the first four entries of $\tilde{V}$. Using the method in Appendix A to find a computed singular value decomposition $\hat{\gamma}\hat{A}\hat{E}\hat{B}$, we have $\|DP - D\hat{\gamma}\hat{A}\hat{E}\hat{B}\|_\infty = O(\delta\|DP\|_\infty)$. Since $\|DP\|_\infty \le \|\tilde{V}\|_\infty$, we have an $O(\|\tilde{V}\|_\infty B_{xy}\delta)$ bound on the infinity norm of (24).

The difference between (24) and the residual $\mathcal{R}$ in (25) is due to the factor of $\hat{\gamma}$, the errors of $O(\delta\hat{\rho}/L)$ in $\hat{\rho}/L$, and the errors between $\hat{R}_i$ and $\hat{S}_i$ and the computed values of $\hat{a}_1 X_i + \hat{a}_2 Y_i$ and $\hat{a}_1 Y_i - \hat{a}_2 X_i$. The additions to the residual norm due to these errors are $O(\lambda\hat{\rho}B_{xy}\delta/L)$ for $\hat{\rho}/L$, $O(B_x\delta + \hat{a}_2 B_{xy}\delta)$ for $\hat{R}_i$, and $O(\hat{\epsilon}B_{xy}\delta)$ for $\hat{S}_i$. Thus repeating the original calculation using $\hat{R}_i$ and $\hat{S}_i$ for the new X_i and Y_i yields a residual with infinity norm

$$O\left(\frac{\|\tilde{V}\|_\infty B_{xy}\delta}{\hat{\gamma}} + \frac{\lambda\hat{\rho}B_{xy}\delta}{L} + B_x\delta + \hat{a}_2 B_{xy}\delta + |\hat{\epsilon}|\, B_{xy}\delta\right). \tag{32}$$

Next we need a good upper bound on the rotation parameter $\hat{a}_2$ that appears in (32). According to (22), the bounding box of $(0,0)$, (R_1, S_1), (R_2, S_2), and (R_3, S_3) has aspect ratio $O(\hat{\epsilon}_0 + \lambda\hat{\rho}_0/L + \lambda\delta)$ before the iterative improvement step, and $O(\hat{\epsilon} + \lambda\hat{\rho}/L + \lambda\delta)$ after rotating by $\sin\hat{a}_2$, hence

$$\hat{a}_2 = O\left(\hat{\epsilon} + \frac{\lambda\hat{\rho}}{L} + \hat{\epsilon}_0 + \frac{\lambda\hat{\rho}_0}{L} + \lambda\delta\right). \tag{33}$$

To complete the proof, we need a bound on the ratio $\|\tilde{V}\|_\infty/\hat{\gamma}$ in (32). Since substituting $\hat{\gamma}\hat{A}\hat{E}\hat{B}$ for $\tilde{P}$ in $\tilde{V}$ affects $\|\tilde{V}\|_\infty$ by a factor of $1 + O(\delta)$, we can use

$$\frac{\tilde{V}}{\hat{\gamma}} \approx \begin{pmatrix} (\hat{a}_1\hat{b}_1 + \hat{\epsilon}\hat{a}_2\hat{b}_2)/\alpha \\ \hat{a}_2\hat{b}_1 - \hat{\epsilon}\hat{a}_1\hat{b}_2 \\ (\hat{a}_1\hat{b}_2 - \hat{\epsilon}\hat{a}_2\hat{b}_1)/\alpha \\ \hat{a}_2\hat{b}_2 + \hat{\epsilon}\hat{a}_1\hat{b}_1 \\ \hat{\rho}/L \end{pmatrix}$$

to estimate $\|\tilde{V}\|_\infty/\hat{\gamma}$. Since

$$\frac{1}{\alpha} = \frac{B_x}{B_{xy}} = O\left(\hat{\epsilon}_0 + \frac{\lambda\hat{\rho}_0}{L} + \lambda\delta\right),$$

we can let

$$\frac{\|\tilde{V}\|_\infty}{\hat{\gamma}} = O\left(\hat{\epsilon} + \frac{\lambda\hat{\rho}}{L} + \hat{\epsilon}_0 + \frac{\lambda\hat{\rho}_0}{L} + \lambda\delta\right)$$

in (32). Substituting this, (22), and (33) into (32) yields the bound on $\|\mathcal{R}\|_\infty$ given in the lemma. □

This suggests the following iterative improvement scheme: First, perform the initial stage as analyzed in Section 3.1. Then if $|\hat{\epsilon}| + \lambda\,|\hat{\rho}|/L \ll 1$, perform an iterative improvement step. In the unlikely event that the new value of $|\hat{\epsilon}| + \lambda\,|\hat{\rho}|/L$ turns out to be much smaller than than the old, perform another step, and proceed until $|\hat{\epsilon}| + \lambda\,|\hat{\rho}|/L$ does not decrease too much or becomes on the order of δ. (This is easily implemented by just a single statement that examines ρ and ϵ and decides whether to start over with the rotated versions of (X_i, Y_i) for $i = 1, 2, 3$.)

After performing iterative improvement steps until $|\hat{\epsilon}| + \lambda\,|\hat{\rho}|/L$ stabilizes, Lemma 3.8 gives

$$\|\mathcal{R}\|_\infty = O\left(\left(\hat{\epsilon} + \frac{\lambda\hat{\rho}}{L} + \lambda\delta\right) B_{xy}\delta\right). \tag{34}$$

Substituting this into (29) gives the bounds on Δr and Δs required by Lemma 3.3.

In practice, a single iterative improvement step usually suffices to reduce the residual to on the order of $|\epsilon| + |\rho|/L$ times its former value, and this makes it essentially zero to within the precision (30). For example, consider a flattened and rotated version of Figure 5a with Bézier control points approximately $(0,0,1)$, $(-188,251,1)$, $(1.6,1.2,1)$, and $(-187,253,1)$. Working in 32-bit floating point with machine precision $\delta = 2^{-24}$, this produced $\hat{\epsilon} \approx 3.8 \times 10^{-8}$ and $\hat{\rho}/L \approx 0.0069$. One iterative improvement step reduced the residual from roughly $4.8\delta B_{xy}$ to $0.013\delta B_{xy}$. This improved the agreement between the given curve and the curve represented by the computed implicit form by a factor of 82.

3.3. Final Stage, Double Point Far Away

This section proves Lemma 3.4 by limiting the error in $G(r,s)$ and using a lower bound on the gradient of G to limit the error in the computed implicit curve $G(r,s) = 0$. Since such a lower bound is not possible when the curve passes through a double point, the lemma requires the ratio of the computed $|\hat{\rho}|/L$ to $|\hat{\epsilon}|$ to be large. The intuitive basis for this is that $\sqrt{r_c^2 + s_c^2}$ should be large compared to L and it is clear from (19) that this can only happen when ρ/ϵ is large compared to L.

When ϵ is small compared to ρ/L, all coefficients of $G(r,s)$ may be found via equations (15) from Section 2.2. This has the advantage that it does not use any of the Z_i parameters that had to be perturbed in Lemma 3.7 when $|\hat{\epsilon}| \le |\hat{\rho}|/L$. Thus we need only consider errors in $G(r,s)$ due to rounding error in evaluating (15).

Lemma 3.9 *Evaluating (15) with floating point of precision δ induces $O(\lambda\rho^3 B_{xy}\delta)$ error in $G(r,s)$ when $|\hat{\epsilon}| \le |\hat{\rho}|/L$, $|s| \le L$ and $|r| \le \hat{B}_r/h_{\min}$.*

Proof. The round-off error in evaluating any coefficient c_{ij} is going to be at most some constant $\delta_1 = O(\delta)$ times the result of evaluating the appropriate right-hand side in (15) with each variable replaced by its absolute value. To get a rough upper bound, replace $|\hat{b}_1|$ and $|\hat{b}_2|$ by one, replace each $|\hat{S}_i|$ by $\hat{B}_{xy}$, and replace each $|\hat{R}_i|$ by $\hat{B}_r$. Using

$$\alpha = \frac{L\hat{B}_r}{|\hat{\rho}|\,B_{xy}} = O\left(\frac{|\hat{\epsilon}|\,L + \lambda\,|\hat{\rho}| + \lambda\delta}{|\hat{\rho}|}\right) = O(\lambda).$$

we get the following upper bounds:

$$\begin{aligned}
&|\Delta c_{30}| \le 7B_{xy}\delta_1, && |\Delta c_{20}| \le 9\,|\hat{\rho}|\,B_{xy}\delta_1,\\
&|\Delta c_{21}| \le (7\alpha\,|\hat{\rho}|/L + 14\,|\hat{\epsilon}|)B_{xy}\delta_1, && |\Delta c_{11}| \le 9\,|\hat{\rho}|\,(\alpha\,|\hat{\rho}|/L + |\hat{\epsilon}|)B_{xy}\delta_1,\\
&|\Delta c_{12}| \le |\hat{\epsilon}|\,(14\alpha\,|\hat{\rho}|/L + 7\,|\hat{\epsilon}|)B_{xy}\delta_1, && |\Delta c_{02}| \le 9\alpha\hat{\rho}^2\,|\hat{\epsilon}|\,B_{xy}\delta_1/L,\\
&|\Delta c_{03}| \le 7\alpha\,|\hat{\rho}|\,\hat{\epsilon}^2 B_{xy}\delta_1/L, && |\Delta c_{10}| \le 3\hat{\rho}^2 B_{xy}\delta_1,\\
& && |\Delta c_{01}| \le 3\alpha\,|\hat{\rho}|^3\,B_{xy}\delta_1/L.
\end{aligned}$$

The total effect on $G(r,s)$ of errors in its coefficients is the sum of all $r^i s^j \Delta c_{ij}$ for $1 \le i+j \le 3$. Since $|r| \le \hat{B}_r/h_{\min} = \alpha\,|\hat{\rho}|$, the total error is at most

$$\left((14\alpha^3 + 18\alpha^2 + 6\alpha)\frac{|\hat{\rho}|^3}{L^3} + (28\alpha^2 + 18\alpha)\frac{\hat{\rho}^2\,|\hat{\epsilon}|}{L^2} + 14\alpha\frac{|\hat{\rho}|\,\hat{\epsilon}^2}{L}\right) L^3 B_{xy}\delta_1.$$

This is $O(\lambda\rho^3 B_{xy}\delta)$ as required. □

To get a first order estimate of the amount by which this error causes the curve $G(r,s) = 0$ to shift, we just divide by the magnitude of the gradient of G at the point (r,s) as given by Theorem C.6 in Appendix C. That theorem shows that if $G(r,s)$ is the function obtained by evaluating (15) exactly using the perturbed curve and $\hat{b}_1$, $\hat{b}_2$, $\hat{\epsilon}$ and $\hat{\rho}$, then under the assumptions of Lemma 3.4, the magnitude of the gradient of G at any point on the perturbed curve is at least

$$\frac{0.031^2\,|\rho|^3\,h_{\min}(1+O(\delta))}{(\lambda-0.43)^2}. \tag{35}$$

Dividing this into the $O(\lambda^3\,|\rho|^3\,B_{xy}\delta)$ bound on the error in $G(r,s)$ gives an estimate of

$$O\left(\frac{\lambda^5 B_{xy}\delta}{h_{\min}}\right) = O(\lambda^5 L\delta). \tag{36}$$

for the shift in the curve $G(r,s)=0$.

We have used a tangent line approximation to determine how far we must go along a line perpendicular to the perturbed curve $G(r,s)=0$ before $G(r,s)$ reaches the error bound from Lemma 3.9. Since this error bound approaches zero as δ approaches zero, the estimate must be valid to within a constant factor for sufficiently small δ. Hence (36) completes the proof of Lemma 3.4.

One way to quantify the idea of a "sufficiently small" δ would be to get an upper bound $\bar{G}_2$ on the second derivative of $G(r,s)$ along the line perpendicular to the curve $G(r,s)=0$. If the first derivative is G_1 at $G(r,s)=0$, it cannot fall to zero within a distance less than $G_1/\bar{G}_2$ and $|G(r,s)|$ must be at least $\frac{1}{2}G_1^2/\bar{G}_2$ at this point. Hence we need only require the bound from Lemma 3.9 to be at most this much.

It would be possible to use an argument very similar to the proof of Lemma 3.9 to get bounds on $\frac{\partial^2 G}{\partial r^2}$, $\frac{\partial^2 G}{\partial r\partial s}$, and $\frac{\partial^2 G}{\partial s^2}$, thereby getting bounds on $\bar{G}_2$ as a function of the curve direction. We could then analyze the proofs of Lemma C.5 and Theorem C.6 in Appendix C to get a stronger bound on G_1 that depends on the curve direction. A preliminary analysis of this type indicates that δ is "sufficiently small" when it is less than some constant times λ^{-6}. For brevity, we leave this as a conjecture.

3.4. Final Stage, Double Point Nearby

The purpose of this section is to prove Lemma 3.5 by doing backward error analysis from the double point centered form. It deals with the case where $\hat{\rho}/\hat{\epsilon}$ is no more than a constant times λL so that (19) makes the distance to the double point obey a similar bound.

Thus we are dealing with the case where the implicit form is the double-point-centered form given by (20) and (21), where (18) is used to evaluate c_{30}, c_{21}, c_{12}, and c_{03}, and (15) is used to evaluate c_{20}, c_{11}, and c_{02}. Section 2.3 suggested eliminating a factor ρ from each of the these equations but it simplifies this discussion to assume that we evaluate the expressions as given to yield computed results $\hat{c}'_{mn}$ for $2\le m+n\le 3$.

Lemma 3.10 *If the coefficients c'_{mn} of the double point centered form are computed as in Section 2.4 from $\hat{\rho}$, $\hat{\epsilon}$, $\hat{b}_1$, $\hat{b}_2$ and the corresponding Beziér control points R_i, S_i, Z_i with errors $\Delta c'_{mn}$, then there is a parametric form for the computed coefficients $\hat{c}'_{mn}$ with the following error bounds: Z_1, Z_2, and Z_3 need to be changed by at most $\|\Delta C_h\|_\infty$, each R_i needs to be changed by at most $\|\Delta C_\theta\|_\infty$, and each S_i needs to be changed by at most $\|\Delta C_\phi\|_\infty$, where*

$$\begin{aligned}
\Delta C_h &= \hat{\rho}^{-1}\hat{\epsilon}(\;\Delta c'_{30}\quad \hat{\epsilon}^{-1}\Delta c'_{21}\quad \hat{\epsilon}^{-2}\Delta c'_{12}\quad \hat{\epsilon}^{-3}\Delta c'_{03}\;),\\
\Delta C_\theta &= \hat{\rho}^{-1}\hat{\epsilon}(\;0\quad \Delta c'_{20}\quad \hat{\epsilon}^{-1}\Delta c'_{11}\quad \hat{\epsilon}^{-2}\Delta c'_{02}\;),\\
\Delta C_\phi &= \hat{\rho}^{-1}(\;\Delta c'_{20}\quad \hat{\epsilon}^{-1}\Delta c'_{11}\quad \hat{\epsilon}^{-2}\Delta c'_{02}\quad 0\;).
\end{aligned}$$

Proof. Let us do backward error analysis on the double-point-centered form

$$\sum_{2\le m+n\le 3} \hat{c}'_{mn}(r-r_c)^m(s-s_c)^n = 0,$$

where $r_c = \hat\rho\hat b_1$ and $s_c = -\hat\rho\hat b_2/\hat\epsilon$ are the true coordinates of the double point. To find a corresponding parametric form parameterized in such a way as to try to match the original $h(t)$, $\theta(t)$, and $\phi(t)$, let

$$\begin{aligned} r - \hat r_c &= \hat b_2\hat\tau - \hat b_1\hat u \\ s - \hat s_c &= \hat\epsilon^{-1}(\hat b_1\hat\tau + \hat b_2\hat u), \end{aligned} \tag{37}$$

where

$$\hat\tau = \frac{r\hat b_2 + s\hat\epsilon\hat b_1}{\hat b_1^2 + \hat b_2^2} \quad \text{and} \quad \hat u = \frac{-r\hat b_1 + s\hat\epsilon\hat b_2}{\hat b_1^2 + \hat b_2^2} + \hat\rho.$$

Substituting (37) into the double-point-centered form, it is not hard to see that

$$\begin{aligned} \sum_{m+n=3} & \hat c'_{mn}\hat\epsilon^{-n}(\hat b_2\hat\tau - \hat b_1\hat u)^m(\hat b_1\hat\tau + \hat b_2\hat u)^n \\ &= -\sum_{m+n=2} \frac{\hat c'_{mn}\hat\epsilon^{-n}(\hat b_2\hat\tau - \hat b_1\hat u)^{m+1}(\hat b_1\hat\tau + \hat b_2\hat u)^n}{r - \hat r_c} \\ &= -\sum_{m+n=2} \frac{\hat c'_{mn}\hat\epsilon^{-n}(\hat b_2\hat\tau - \hat b_1\hat u)^m(\hat b_1\hat\tau + \hat b_2\hat u)^{n+1}}{\hat\epsilon(s - \hat s_c)}. \end{aligned} \tag{38}$$

This yields expressions for $r - r_c$ and $s - s_c$ that can be written parametrically by letting the ratio $t : 1-t$ be the same as $\hat\tau : \hat u$. This yields $r - r_c = \hat\theta(t)/\hat h(t)$ and $s - s_c = \hat\phi(t)/\hat h(t)$, where

$$\begin{aligned} \hat h(t) &= \nu \sum_{m+n=3} \hat c'_{mn}\hat\epsilon^{-n}(\hat b_2 t - \hat b_1(1-t))^m(\hat b_1 t + \hat b_2(1-t))^n, \\ \hat\theta(t) &= -\nu \sum_{m+n=2} \hat c'_{mn}\hat\epsilon^{-n}(\hat b_2 t - \hat b_1(1-t))^{m+1}(\hat b_1 t + \hat b_2(1-t))^n, \\ \hat\phi(t) &= -\nu\hat\epsilon^{-1} \sum_{m+n=2} \hat c'_{mn}\hat\epsilon^{-n}(\hat b_2 t - \hat b_1(1-t))^m(\hat b_1 t + \hat b_2(1-t))^{n+1}, \end{aligned} \tag{39}$$

and ν is a constant factor that needs to be chosen in order to try to match the original $h(t)$, $\theta(t)$, and $\phi(t)$.

We want to choose ν so that $\hat h(t)$ would match $h(t)$ if all $\hat c'_{mn}$ were exact. Since we observed in Section 2.3 that the cubic terms of $G(r,s)$ are identical to those of $\rho\epsilon^{-1}h(\tau,u)$, the cubic terms in (38) should be like those of $\hat\rho\hat\epsilon^{-1}h(\hat\tau,\hat u)$. Thus substituting t for $\hat\tau$ and $1-t$ for $\hat u$ would yield $\hat\rho\hat\epsilon^{-1}h(t)$ for the part of (38) that becomes $\hat h(t)$. Thus letting

$$\nu = \hat\rho^{-1}\hat\epsilon,$$

makes $\hat h(t)$, $\hat\theta(t)$, and $\hat\phi(t)$ correct except for the error in the $\hat c'_{mn}$.

Expanding (39) to make the Bézier control points apparent yields $\hat h(t) = \hat C_h Q T_t$, $\hat\theta(t) = \hat C_\theta Q T_t$, $\hat\phi(t) = \hat C_\phi Q T_t$, where

$$Q = \begin{pmatrix} -\hat b_1^3 & \hat b_1^2\hat b_2 & -\hat b_1\hat b_2^2 & \hat b_2^3 \\ \hat b_1^2\hat b_2 & \frac13(\hat b_1^3 - 2\hat b_1\hat b_2^2) & \frac13(\hat b_2^3 - 2\hat b_1^2\hat b_2) & \hat b_1\hat b_2^2 \\ -\hat b_1\hat b_2^2 & \frac13(\hat b_1^2\hat b_2 - 2\hat b_2^3) & \frac13(2\hat b_1\hat b_2^2 - \hat b_1^3) & \hat b_1^2\hat b_2 \\ \hat b_2^3 & \hat b_1\hat b_2^2 & \hat b_1^2\hat b_2 & \hat b_1^3 \end{pmatrix} \quad \text{and} \quad T_t = \begin{pmatrix} (1-t)^3 \\ 3t(1-t)^2 \\ 3t^2(1-t) \\ t^3 \end{pmatrix}$$

and

$$\begin{aligned} \hat C_h &= \hat\rho^{-1}\hat\epsilon(\ \hat c'_{30} \quad \hat\epsilon\hat c'_{21} \quad \hat\epsilon^{-2}\hat c'_{12} \quad \hat\epsilon^{-3}\hat c'_{03}\ ), \\ \hat C_\theta &= \hat\rho^{-1}\hat\epsilon(\ 0 \quad \hat c'_{20} \quad \hat\epsilon^{-1}\hat c'_{11} \quad \hat\epsilon^{-2}\hat c'_{02}\ ), \\ \hat C_\phi &= \hat\rho^{-1}(\ \hat c'_{20} \quad \hat\epsilon^{-1}\hat c'_{11} \quad \hat\epsilon^{-2}\hat c'_{02} \quad 0\ ). \end{aligned}$$

Thus the Bézier control points are given by the elements of $\hat C_\theta Q$, $\hat C_\phi Q$, and $\hat C_h Q$, and the lemma follows from the fact that $\|Q\|_\infty \le 1 + O(\delta)$. $\square$

To apply Lemma 3.10, we need bounds on the errors in $\hat{C}_h$, $\hat{C}_\theta$, and $\hat{C}_\phi$ due to round-off error and the use of Z_i and the computed values $\hat{R}_i$, $\hat{S}_i$, instead of the perturbed versions. These bounds depend on whether iterative improvement is used to reduce the bound on $\|\mathcal{R}\|_\infty$ from (23), but we can always use Lemma 3.7 to limit the perturbation as follows:

$$|\Delta R_i| \leq \frac{11}{10} \|\mathcal{R}\|_\infty, \qquad |\Delta S_i| \leq \frac{11 \|\mathcal{R}\|_\infty}{10\,|\hat{\epsilon}|}, \quad \text{and} \quad |Z_i| \leq \frac{\|\mathcal{R}\|_\infty}{\max(|\hat{\rho}|\,, |\hat{\epsilon}|\,L)}. \tag{40}$$

Hence the following lemma gives bounds on the $\|\Delta C_\theta\|_\infty$, $\|\Delta C_\phi\|_\infty$, and $\|\Delta C_h\|_\infty$ needed in Lemma 3.10:

Lemma 3.11 *Consider the errors $\Delta c'_{mn}$ due to rounding error in (15), (18), and (21) and perturbations with upper bounds*

$$|\Delta R_i| \leq \alpha_1, \qquad |\Delta S_i| \leq \frac{\alpha_1}{|\hat{\epsilon}|}, \quad \text{and} \quad |\Delta Z_i| \leq \frac{\alpha_1}{\max(|\hat{\rho}|\,, |\hat{\epsilon}|\,L/\lambda)}.$$

Using these errors in ΔC_h, ΔC_θ, and ΔC_ϕ makes

$$\|\Delta C_h\|_\infty \leq \frac{24\alpha_2}{\max(|\hat{\rho}|\,, |\hat{\epsilon}|\,L/\lambda)}, \qquad \|\Delta C_\theta\|_\infty \leq 42\lambda\alpha_2, \quad \text{and} \quad \|\Delta C_\phi\|_\infty \leq \frac{42\lambda\alpha_2}{|\hat{\epsilon}|},$$

where α_2 is a constant times $\alpha_1 + (|\hat{\epsilon}| + \lambda\,|\hat{\rho}|\,/L + \lambda\delta)B_{xy}\delta$.

Proof. For convenience, let

$$\beta = \max\left(|\hat{\rho}|\,, \frac{|\hat{\epsilon}|\,L}{\lambda}\right).$$

Since (15), (18), and (21) are linear in R_i, S_i, and Z_i, the error due to the perturbation is at most the result of evaluating (20) and (21) with R_i replaced by α_1, S_i replaced by $\alpha_1/\,|\hat{\epsilon}|$, Z_i replaced by α_1/β, and each other variable replaced by a bound on its absolute value. Of course it is necessary to use (18) for c_{30}, c_{21}, c_{12}, and c_{03} in (20) and (15) for c_{20}, c_{11}, and c_{02} in (21). Using $|\hat{b}_1| \leq 1$ and $|\hat{b}_2| \leq 1$ gives error bounds

$$\begin{aligned}
&|\Delta c'_{30}| \leq 8\,|\hat{\epsilon}|^{-1}\,|\hat{\rho}|\,\alpha_1/\beta && \\
&|\Delta c'_{21}| \leq 24\,|\hat{\rho}|\,\alpha_1/\beta && |\Delta c'_{20}| \leq (9\,|\hat{\rho}|\,|\hat{\epsilon}|^{-1} + 12\hat{\rho}^2\,|\hat{\epsilon}|^{-1}/\beta)\alpha_1, \\
&|\Delta c'_{12}| \leq 24\,|\hat{\epsilon}|\,|\hat{\rho}|\,\alpha_1/\beta && |\Delta c'_{11}| \leq (18\,|\hat{\rho}| + 24\hat{\rho}^2/\beta)\alpha_1, \\
&|\Delta c'_{03}| \leq 8\hat{\epsilon}^2\,|\hat{\rho}|\,\alpha_1/\beta && |\Delta c'_{02}| \leq (9\,|\hat{\rho}|\,|\hat{\epsilon}| + 12\hat{\rho}^2\,|\hat{\epsilon}|\,/\beta)\alpha_1.
\end{aligned} \tag{41}$$

Since the round-off error in evaluating a sum of a constant number of terms is $\delta_1 = O(\delta)$ times the sum of the absolute values, the effect of round-off error can be included by making the following changes in the derivation of the bounds on $|\Delta c'_{mn}|$: the replacement for R_i becomes

$$\alpha_1 + \hat{B}_r\delta_1; \tag{42}$$

the replacement for S_i becomes

$$\frac{\alpha_1}{|\hat{\epsilon}|} + B_{xy}\delta_1; \tag{43}$$

and the replacement for Z_i becomes

$$\frac{\alpha_1}{\beta} + \lambda h_{\min}\delta_1. \tag{44}$$

By choosing α_2 so that (42) is at most α_2, (43) is at most $\alpha_2/\,|\hat{\epsilon}|$, and (44) is at most α_2/β, we can account for round-off error by replacing α_1 with α_2 in (41). In fact the lemma does allow α_2 to be this large since

$$\alpha_1 + \lambda\beta h_{\min}\delta_1 = \alpha_1 + \frac{\lambda B_{xy}\delta_1}{L}\max\left(|\hat{\rho}|\,, \frac{|\hat{\epsilon}|\,L}{\lambda}\right) = \alpha_1 + O\left(\left(\hat{\epsilon} + \frac{\lambda\hat{\rho}}{L}\right)B_{xy}\delta\right),$$

and because of (22),

$$\alpha_1 + \hat{B}_r \delta_1 = \alpha_1 + O((\hat{\epsilon} + \lambda\hat{\rho}/L + \lambda\delta) B_{xy}\delta).$$

Replacing α_1 with α_2 in (41) yields

$$\begin{aligned}
\|\Delta C_h\|_\infty &\le \max\left(\frac{8\alpha_2}{\beta},\ \frac{24\alpha_2}{\beta},\ \frac{24\alpha_2}{\beta},\ \frac{8\alpha_2}{\beta}\right), \\
\|\Delta C_\theta\|_\infty &\le \max\left(0,\ \left(9 + 12\frac{|\hat{\rho}|}{\beta}\right)\alpha_2,\ \left(18 + 24\frac{|\hat{\rho}|}{\beta}\right)\alpha_2,\ \left(9 + 12\frac{|\hat{\rho}|}{\beta}\right)\alpha_2\right), \\
\|\Delta C_\phi\|_\infty &\le \max\left(\left(9 + 12\frac{|\hat{\rho}|}{|\hat{\epsilon}|\,\beta}\right)\alpha_2,\ \left(18 + 24\frac{|\hat{\rho}|}{|\hat{\epsilon}|\,\beta}\right)\alpha_2,\ \left(9 + 12\frac{|\hat{\rho}|}{|\hat{\epsilon}|\,\beta}\right)\alpha_2,\ 0\right),
\end{aligned}$$

and since $|\hat{\rho}|/\beta \le 1$ the lemma follows. □

We are now ready to use the above lemmas to prove Lemma 3.5 as stated in the introductory part of Section 3. The perturbations (40) satisfy the assumptions of Lemma 3.11 if we let $\alpha_1 = \frac{11}{10}\|\mathcal{R}\|_\infty$. Without iterative improvement, Lemma 3.6 gives $\|\mathcal{R}\|_\infty = O(\lambda B_{xy}\delta)$, hence $\alpha_2 = O(\lambda B_{xy}\delta)$ in Lemma 3.11. Since the errors $\Delta\theta$, $\Delta\phi$, and Δh in $\theta(t)$, $\phi(t)$, and $h(t)$ are bounded by the errors in their control points, Lemma 3.10 guarantees

$$|\Delta\theta| \le \|C_\theta\|_\infty, \qquad |\Delta\phi| \le \|C_\phi\|_\infty, \quad \text{and} \quad |\Delta h| \le \|C_h\|_\infty.$$

Thus Lemma 3.11 gives

$$\begin{aligned}
\frac{\Delta h}{h_{\min}} &= O\left(\frac{\lambda B_{xy}\delta}{\max(|\hat{\rho}|, |\hat{\epsilon}|\,L/\lambda) h_{\min}}\right) = O\left(\frac{\lambda^2\delta}{\max(|\hat{\epsilon}|, \lambda\,|\hat{\rho}|/L)}\right) = O\left(\frac{\lambda^2\delta}{\hat{\epsilon}}\right) \\
\frac{\Delta\theta}{h_{\min}} &= O\left(\frac{\lambda^2 B_{xy}\delta}{h_{\min}}\right) = O(\lambda^2 L\delta) \\
\frac{\Delta\phi}{h_{\min}} &= O\left(\frac{\lambda^2 B_{xy}\delta}{\hat{\epsilon} h_{\min}}\right) = O\left(\frac{\lambda^2 L\delta}{\hat{\epsilon}}\right)
\end{aligned}$$

Substituting these equations into

$$\begin{aligned}
|\Delta r| &= \left|\frac{\theta(t) + \Delta\theta}{h(t) + \Delta h} - \frac{\theta(t)}{h(t)}\right| = \left|\frac{h(t)\Delta\theta - \theta(t)\Delta h}{h(t)^2 + h(t)\Delta h}\right| \le \frac{|\Delta\theta|}{h_{\min}} + \frac{\Delta h \hat{B}_r}{h_{\min}^2} \\
|\Delta s| &= \left|\frac{\phi(t) + \Delta\phi}{h(t) + \Delta h} - \frac{\phi(t)}{h(t)}\right| = \left|\frac{h(t)\Delta\phi - \phi(t)\Delta h}{h(t)^2 + h(t)\Delta h}\right| \le \frac{|\Delta\phi|}{h_{\min}} + \frac{\Delta h \hat{B}_{xy}}{h_{\min}^2}
\end{aligned}$$

yields

$$|\Delta r| = O\left(\lambda^2\left(1 + \frac{\lambda\delta}{|\hat{\epsilon}|}\right) L\delta\right) \quad \text{and} \quad |\Delta s| = O\left(\frac{\lambda^2 L\delta}{|\hat{\epsilon}|}\right). \tag{45}$$

as required by Lemma 3.5.

Lemma 3.5 gives stronger bounds on $|\Delta r|$ and $|\Delta s|$ when iterative improvement is used. These could be obtained by repeating the derivation of (45) using (34) from Section 3.2 instead of Lemma 3.6, but it is easier just to observe that the error bounds are proportional to the α_2 parameter in Lemma 3.11. With iterative improvement (34) makes α_1

$$O\left(\left(\hat{\epsilon} + \frac{\lambda\hat{\rho}}{L} + \lambda\delta\right) B_{xy}\delta\right),$$

hence α_2 is also of this order. Thus the bound on α_2 is reduced to

$$O\left(\frac{\hat{\epsilon}}{\lambda} + \frac{\hat{\rho}}{L} + \delta\right)$$

of its former value, and Lemma 3.5's assumption that $|\hat{\rho}| \le 33(\lambda - \frac{1}{9})L\,|\hat{\epsilon}|$ makes this $O(\lambda\hat{\epsilon} + \delta)$. Multiplying the right side of (45) by this factor yields the bounds required by Lemma 3.5 under the iterative improvement scheme of Section 3.2.

4. Conclusion

The computation outlined in Section 2 is unique in that no other work has been done on numerically stable techniques for finding implicit forms for the parametric curves commonly used in graphics and computer typesetting applications. In spite of the difficulty in proving numerical stability and the need to derive alternative implicitization formulas in Sections 2.3 and 2.4, the algorithm itself is not much more complicated than the algorithm by Sederberg that requires exact rational arithmetic [9]. The main difference is the singular value decomposition and the use of rotated coordinates.

We have not discussed running time because everything is in-line code except for the loop added when iterative improvement is used. That loop could probably be removed by using some other method to compute an initial rotation, but this seems totally unnecessary in practice since iterative improvement has never been observed to require more than one extra iteration. In any case, the error bounds without iterative improvement are not too bad if we disallow "flat spots" of the type shown in Figure 5.

The main numerical result is that when the constant λ is not too large, the error between the given parametric curve defined on $0 \le t \le 1$ and the curve represented by the computed implicit form is on the order of the machine precision times the overall dimensions of the parametric curve. The exponent of λ in the error bounds may be overly pessimistic, but the example of Figure 4 near the beginning of Section 3 shows that some dependence on λ is clearly required. Note that we get good error bounds in the important case of a polynomial cubic where $\lambda = 1$ by definition. Even for rational cubics, large values of λ can often be avoided by using the rational reparameterization mentioned at the beginning of Section 2 to force Z_0 to be equal to Z_3.

A. Two by Two Singular Value Decomposition

The singular value decomposition of a matrix P is $P = ADB$ where A and B are orthogonal matrices and D is diagonal. General techniques for finding the singular value decomposition can be found in [3], so we concentrate on finding a simple solution for the two by two case where the residual in the computed solution $\hat{A}\hat{D}\hat{B}$ is small in the sense that $\|D'(P - \hat{A}\hat{D}\hat{B})\| \ll \|D'P\|$ for any diagonal matrix D'.

Letting

$$A = \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix}, \qquad D = \begin{pmatrix} \gamma & 0 \\ 0 & \gamma\epsilon \end{pmatrix}, \qquad B = \begin{pmatrix} b_1 & b_2 \\ -b_2 & b_1 \end{pmatrix},$$

where $|\epsilon| \le 1$, the problem can be viewed as writing P as the sum of two rank-one matrices:

$$\begin{pmatrix} P_{0x} & P_{1x} \\ P_{0y} & P_{1y} \end{pmatrix} = \gamma \begin{pmatrix} a_1 b_1 & a_1 b_2 \\ a_2 b_1 & a_2 b_2 \end{pmatrix} + \gamma\epsilon \begin{pmatrix} a_2 b_2 & -a_2 b_1 \\ -a_1 b_2 & a_1 b_1 \end{pmatrix}.$$

Thus we have

$$\gamma \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix} \begin{pmatrix} a_1 b_1 \\ a_2 b_2 \end{pmatrix} = \begin{pmatrix} P_{0x} \\ P_{1y} \end{pmatrix}, \qquad \gamma \begin{pmatrix} 1 & -\epsilon \\ -\epsilon & 1 \end{pmatrix} \begin{pmatrix} a_2 b_1 \\ a_1 b_2 \end{pmatrix} = \begin{pmatrix} P_{0y} \\ P_{1x} \end{pmatrix}, \tag{46}$$

or alternatively

$$\gamma \begin{pmatrix} 1+\epsilon & 1+\epsilon \\ \epsilon-1 & 1-\epsilon \end{pmatrix} \begin{pmatrix} a_1 b_1 \\ a_2 b_2 \end{pmatrix} = \begin{pmatrix} P_{1y} + P_{0x} \\ P_{1y} - P_{0x} \end{pmatrix},$$
$$\gamma \begin{pmatrix} -1-\epsilon & 1+\epsilon \\ 1-\epsilon & 1-\epsilon \end{pmatrix} \begin{pmatrix} a_2 b_1 \\ a_1 b_2 \end{pmatrix} = \begin{pmatrix} P_{1x} - P_{0y} \\ P_{1x} + P_{0y} \end{pmatrix}. \tag{47}$$

Once ϵ and γ are known, either (46) or (47) can be used to find a_1, a_2, b_1, and b_2. We therefore make use of the fact that the quantities

$$\gamma_0 = \sqrt{(P_{1y}+P_{0x})^2+(P_{1x}-P_{0y})^2} = \gamma(1+\epsilon)\sqrt{(a_1b_1+a_2b_2)^2+(-a_2b_1+a_1b_2)^2},$$
$$\gamma_1 = \sqrt{(P_{1y}-P_{0x})^2+(P_{1x}+P_{0y})^2} = \gamma(1-\epsilon)\sqrt{(-a_1b_1+a_2b_2)^2+(a_2b_1+a_1b_2)^2}$$

are easily evaluated to good relative accuracy. Since $a_1^2+a_2^2=b_1^2+b_2^2=1$ we have $\gamma_0=\gamma(1+\epsilon)$, $\gamma_1=\gamma(1-\epsilon)$, and therefore

$$\gamma = \frac{\gamma_0+\gamma_1}{2}.$$

A good way to evaluate ϵ is to use the fact that $\det|P| = P_{0x}P_{1y}-P_{1x}P_{0y} = \gamma^2\epsilon$ and thus

$$\epsilon = \frac{P_{0x}P_{1y}-P_{1x}P_{0y}}{\gamma^2}.$$

This makes the error in ϵ small compared to

$$\frac{\min(|P_{0x}|+|P_{1x}|,\ |P_{0y}|+|P_{1y}|)}{\|P\|_\infty}. \tag{48}$$

The next task is to solve (46) or (47) for the desired values of a_1b_1, a_1b_2, a_2b_1, and a_2b_2. Using (46) when $|\epsilon|<\frac{1}{2}$ and (47) when $|\epsilon|\geq\frac{1}{2}$ ensures that the matrices are well-conditioned so that solution vectors $(a_1b_1\ a_2b_2)^T$ and $(a_2b_1\ a_1b_2)^T$ are obtained with good accuracy.

To complete the singular value decomposition, set

$$(a_i, a_{1-i}) \leftarrow (V_{i,j}, V_{1-i,j}) \Big/ \sqrt{V_{i,j}^2+V_{1-i,j}^2},$$
$$(b_j, b_{1-j}) \leftarrow (V_{i,j}, V_{i,1-j}) \Big/ \sqrt{V_{i,j}^2+V_{i,1-j}^2},$$

where $V_{i',j'}$ is the desired value of $a_{i'}b_{j'}$ and i and j are chosen so that the unused value $V_{1-i,1-j}$ is the smallest of the four. This can cause $a_{1-i}b_{1-j}$ to deviate from $V_{1-i,1-j}$, but only by an amount on the order of machine precision times V_{ij} times the ratio (48). The final result is that each component of $P-ADB$ is small compared to magnitude of the corresponding component of P plus the ratio (48) times the magnitude of the diagonally opposite component.

B. Proof of the Rotation Theorem

A key idea underlying all the numerical stability arguments is that the (r,s) coordinate system is chosen wisely when ρ/L and ϵ are both small. In that case, the following lemma shows that the curve lies close to the line $r=0$ in the sense that

$$B_r = \max_{i=1,2,3} |R_i|$$

is small compared to

$$B_{xy} = \max_{i=1,2,3} \sqrt{R_i^2+S_i^2}.$$

Lemma B.1 *If ρ, ϵ and all (R_i,S_i) are as obtained in Section 2, then*

$$B_r < \left(11.25\,|\epsilon| + \frac{7.95\lambda\,|\rho|}{L}\right) B_{xy}.$$

Proof. The system

$$\begin{pmatrix} 0 & 0 & 3R_1 & 3S_1 & -LZ_0 \\ 3R_1 & 3S_1 & 3R_2 & 3S_2 & -3LZ_1 \\ 3R_2 & 3S_2 & R_3 & S_3 & -3LZ_2 \\ R_3 & S_3 & 0 & 0 & -LZ_3 \end{pmatrix} \begin{pmatrix} b_1 \\ -b_2\epsilon \\ b_2 \\ b_1\epsilon \\ \rho/L \end{pmatrix} = 0$$

analogous to (9) obtained by equating $\tau^i u^{4-i}$ terms in (16) can be also be written

$$\begin{pmatrix} 3b_2 & 0 & 0 \\ 3b_1 & 3b_2 & 0 \\ 0 & 3b_1 & b_2 \\ 0 & 0 & b_1 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix} = \begin{pmatrix} -3\epsilon b_1 S_1 + \rho Z_0 \\ 3\epsilon b_2 S_1 - 3\epsilon b_1 S_2 + 3\rho Z_1 \\ 3\epsilon b_2 S_2 - \epsilon b_1 S_3 + 3\rho Z_2 \\ \epsilon b_2 S_3 + \rho Z_3 \end{pmatrix}. \tag{49}$$

Thus an expression for $(R_1\ R_2\ R_3)^T$ is obtained by multiplying the right-hand side of (49) by either of the following left inverse matrices:

$$M_1 = \frac{1}{b_2}\begin{pmatrix} \frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3}b_1/b_2 & \frac{1}{3} & 0 & 0 \\ b_1^2/b_2^2 & -b_1/b_2 & 1 & 0 \end{pmatrix}, \qquad M_2 = \frac{1}{b_1}\begin{pmatrix} 0 & \frac{1}{3} & -\frac{1}{3}b_2/b_1 & \frac{1}{3}b_2^2/b_1^2 \\ 0 & 0 & \frac{1}{3} & -\frac{1}{3}b_2/b_1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Of course B_r is the infinity norm of $(R_1\ R_2\ R_3)^T$, and we can get an upper bound on this by multiplying $\|M_1\|_\infty$ or $\|M_2\|_\infty$ by the upper bound

$$\left(3\sqrt{2}\,|\epsilon| + \frac{3\lambda\,|\rho|}{L}\right) B_{xy}$$

on the infinity norm of the right-hand side of (49). Thus we complete the proof by noting that

$$\|M_1\|_\infty \le \sqrt{1+3^{-2/3}}(3^{-2/3}+3^{-1/3}+1) < 2.65$$

when $|b_1/b_2| \le 3^{-1/3}$ and $\|M_2\|_\infty < 2.65$ when $|b_2/b_1| < 3^{1/3}$. □

Theorem B.2 *If replacing R_i, S_i, etc. by approximate values $\hat{R}_i$, $\hat{S}_i$, $\hat{Z}_i$, $\hat{b}_i$, $\hat{\epsilon}$, and $\hat{\rho}$ in the left-hand side of (23) yields a vector $\mathcal{R}$, then*

$$\hat{B}_r < \left(11.25\,|\hat{\epsilon}| + \frac{7.95\lambda\,|\hat{\rho}|}{L}\right) B_{xy} + 2.65\,\|\mathcal{R}\|_\infty$$

where $\hat{B}_r$ is analogous to B_r.

Proof. Everything goes as in the proof of Lemma B.1, except the residual vector $\mathcal{R}$ is subtracted from the right-hand side of (49). □

C. Proof of the Gradient Theorem

In this appendix, we derive a lower bound on the magnitude of the gradient of the function $G(r,s)$ on the curve $\big(r(t), s(t)\big)$ assuming that $|\epsilon|$ is small compared to $|\rho|/L$. Everything is based on the assumption that the equations in Section 2 hold exactly as given. The way to use the results derived here in the presence of numerical error is to apply them to the perturbed curve that agrees with the computed values of ρ, ϵ, b_1, and b_2.

Lemma C.1 *When evaluated on the curve*

$$\big(r(t), s(t)\big) = \left(\frac{\theta(t)}{h(t)}, \frac{\phi(t)}{h(t)}\right), \tag{50}$$

the gradient of the function $G(r,s)$ defined in Section 2.2 is

$$\rho(\tau+u)h(t)\left(-\frac{ds}{dt},\frac{dr}{dt}\right),$$

where $\tau = rb_2+\epsilon sb_1$ and $u=-rb_1+\epsilon sb_2+\rho$.

Proof. The derivation in Section 2.2 shows that $t/(1-t)=\tau/u$ when $r=r(t)$ and $s=s(t)$. Thus we let

$$t=\frac{\tau}{\tau+u}$$

and observe that

$$\begin{aligned}\tau &= b_2(r-r_c)+\epsilon b_1(s-s_c)\\ u &= -b_1(r-r_c)+\epsilon b_2(s-s_c)\end{aligned} \tag{51}$$

when $r_c=\rho b_1$ and $s_c=-\rho b_2/\epsilon$ as given by (19). Hence a rational substitution for t allows

$$\bar{t}=\frac{(b_1-b_2)t+b_2}{(b_1+b_2)t-b_1}=\frac{(b_1-b_2)\tau+b_2(\tau+u)}{(b_1+b_2)\tau-b_1(\tau+u)}=\frac{\epsilon(s-s_c)}{r-r_c}. \tag{52}$$

Substituting this into the result of factoring $r-r_c$ out of (51) yields

$$\tau=(b_2+b_1\bar{t})(r-r_c)\quad\text{and}\quad u=(-b_1+b_2\bar{t})(r-r_c).$$

This allows $G(r,s)$ to be written in terms of polynomial functions $\bar{h}$, $\bar{\theta}$, and $\bar{\phi}$, where

$$\bar{h}(\bar{t})=\frac{h(t)}{((b_1+b_2)t-b_1)^3},\quad \bar{\theta}(\bar{t})=\frac{\theta(t)}{((b_1+b_2)t-b_1)^3},\quad \bar{\phi}(\bar{t})=\frac{\phi(t)}{((b_1+b_2)t-b_1)^3}.$$

Since

$$\begin{aligned}(b_1+b_2)\bar{t}+b_2-b_1 &= \frac{(b_1+b_2)((b_1-b_2)t+b_2)+(b_2-b_1)((b_1+b_2)t-b_1)}{(b_1+b_2)t-b_1}\\ &= \frac{1}{(b_1+b_2)t-b_1},\end{aligned}$$

we can write

$$h(\tau,u)=(\tau+u)^3h(t)=((b_1+b_2)\bar{t}+b_2-b_1)^3(r-r_c)^3h(t)=(r-r_c)^3\bar{h}(\bar{t}), \tag{53}$$

and similarly for $\theta(\tau,u)$ and $\phi(\tau,u)$. Hence the right-hand side of (17) becomes

$$G(r,s)=\frac{\rho}{\epsilon}(r-r_c)^3\left(\bar{h}(\bar{t})-\frac{b_2\bar{\theta}(\bar{t})+\epsilon b_1\bar{\phi}(\bar{t})}{rb_2+\epsilon sb_1}\right).$$

The slightly simpler form of $G(r,s)$ results from noting that (52) can be written

$$\bar{t}=\epsilon\frac{\bar{\phi}(\bar{t})-s_c\bar{h}(\bar{t})}{\bar{\theta}(\bar{t})-r_c\bar{h}(\bar{t}),}$$

and therefore $\bar{\theta}(\bar{t})-r_c\bar{h}(\bar{t})$ is a quadratic polynomial in $\bar{t}$. Thus the polynomial

$$\frac{\rho}{\epsilon}(r-r_c)^2\big((r-r_c)\bar{h}(\bar{t})+r_c\bar{h}(\bar{t})-\bar{\theta}(\bar{t})\big). \tag{54}$$

has cubic terms identical to $G(r,s)$ and is clearly zero on the curve (50). Since (50) cannot be expressed implicitly in terms of a quadratic polynomial, (54) must be equal to $G(r,s)$. (The same result can be obtained by using (19) and (17) to expand (54) and $G(r,s)$).

Using

$$\frac{d\bar{t}}{dr} = \frac{d}{dr}\left(\frac{\epsilon(s-s_c)}{r-r_c}\right) = -\frac{\epsilon(s-s_c)}{(r-r_c)^2} = -\frac{\bar{t}}{r-r_c}$$

and $\frac{d\bar{t}}{ds} = \epsilon/(r-r_c)$, the derivatives of (54) become

$$\frac{\partial G}{\partial r} = 2\frac{G(r,s)}{r-r_c} + \frac{\rho}{\epsilon}(r-r_c)^2\left(\bar{h}(\bar{t}) - \bar{t}\bar{h}'(\bar{t}) - \frac{r_c\bar{t}\bar{h}'(\bar{t})}{r-r_c} + \frac{\bar{t}\bar{\theta}'(\bar{t})}{r-r_c}\right),$$
$$\frac{\partial G}{\partial s} = \frac{\rho}{\epsilon}(r-r_c)^2\left(\frac{\epsilon r\bar{h}'(\bar{t})}{r-r_c} - \frac{\epsilon\bar{\theta}'(\bar{t})}{r-r_c}\right).$$

Since differentiating (52) yields

$$\frac{d\bar{t}}{dt} = -\frac{b_1^2+b_2^2}{((b_1+b_2)t-b_1)^2} = -\frac{(\tau+u)^2}{\big((b_1+b_2)\tau - b_1(\tau+u)\big)^2} = -\frac{(\tau+u)^2}{(r-r_c)^2},$$

we have

$$\begin{aligned}
\frac{ds}{dt} &= \frac{d}{dt}(s-s_c)\\
&= \frac{d}{dt}\frac{\bar{t}}{\epsilon}\left(\frac{\bar{\theta}(\bar{t})}{\bar{h}(\bar{t})} - r_c\right)\\
&= -\frac{(\tau+u)^2}{\epsilon(r-r_c)^2}\left(\frac{\bar{\theta}(\bar{t})}{\bar{h}(\bar{t})} - r_c + \frac{\bar{t}\bar{\theta}'(\bar{t})}{\bar{h}(\bar{t})} - \frac{\bar{t}\bar{\theta}(\bar{t})\bar{h}'(\bar{t})}{\bar{h}^2(\bar{t})}\right),\\
\frac{dr}{dt} &= \frac{d}{dt}\left(-\frac{\bar{\theta}(\bar{t})}{\bar{h}(\bar{t})}\right)\\
&= -\frac{(\tau+u)^2}{(r-r_c)^2}\left(\frac{\bar{\theta}'(\bar{t})}{\bar{h}(\bar{t})} - \frac{\bar{\theta}(\bar{t})\bar{h}'(\bar{t})}{\bar{h}^2(\bar{t})}\right).
\end{aligned}$$

On the curve, we are allowed to substitute zero for $G(r,s)$ in the expression for $\frac{\partial G}{\partial r}$ and r for $\bar{\theta}(\bar{t})/\bar{h}(\bar{t})$ in the expressions for $\frac{ds}{dt}$ and $\frac{dr}{dt}$, giving

$$\left(\frac{\partial G}{\partial r}, \frac{\partial G}{\partial s}\right) = \frac{\rho(r-r_c)^3\bar{h}(\bar{t})}{(\tau+u)^2}\left(-\frac{ds}{dt}, \frac{dr}{dt}\right).$$

From (53) we have $(r-r_c)^3\bar{h}(\bar{t}) = (\tau+u)^3 h(t)$, hence the lemma follows. □

The lemma gives the gradient as a product of four factors, and we need lower bounds on each of them that hold when ρ/L is large compared to ϵ. The $\tau+u$ factor is of particular interest because (51) shows that a nonzero lower bound ensures that the curve does not pass through the double point. The lower bound comes from two lemmas: the first shows that $|\tau+u| \geq |b_2\rho| - |\epsilon|\,L$, and the second gives a lower bound on $|b_2|$ that depends on λ.

Lemma C.2 *When $b_1^2+b_2^2=1$, $\tau = rb_2+\epsilon sb_1$, $u = -rb_1+\epsilon sb_2+\rho$, and $0 \leq \tau/(\tau+u) \leq 1$, the bound*

$$|\tau+u| \geq \frac{|b_2\rho+\epsilon s|}{\max(|b_1|,|b_2|)}$$

is satisfied.

Proof. The condition on $\tau/(\tau+u)$ allows us to restrict our attention to the portion of the (r,s) plane where τ and u do not have opposite signs. For fixed s, the set of (r,s) values under consideration form a line segment or one or two semi-infinite rays. In any case, the minimum value of $|\tau+u|$ occurs at the intersection with one of the boundary lines $\tau=0$ or $u=0$. Thus the minimum occurs when

$$r = -\frac{b_1\epsilon s}{b_2} \quad \text{or} \quad r = \frac{\rho+b_2\epsilon s}{b_1}.$$

(If $b_1 = 0$ or $b_2 = 0$ then there is only one intersection point and only one of the choices for r is viable).

Using the suggested values for r in the expressions for τ and u yields

$$\tau + u = \frac{(b_1^2 + b_2^2)\epsilon s + b_2\rho}{b_2} \quad \text{or} \quad \tau + u = \frac{(b_1^2 + b_2^2)\epsilon s + b_2\rho}{b_1},$$

hence the lemma follows. □

Lemma C.2 provides the necessary lower bound for $\tau + u$, but the bound is only useful if we can guarantee that b_2 is not too small. The necessary lemma is best expressed in rotated coordinates $(\epsilon\xi, \eta)$, where

$$\begin{aligned}
\xi &= \frac{b_2 - b_1}{\rho} r + \frac{\epsilon(b_1 + b_2)}{\rho} s + 1, \\
\eta &= \frac{\epsilon^2(b_1 + b_2)}{\rho} r + \frac{\epsilon(b_1 - b_2)}{\rho} s + \bar{t}_0, \\
\bar{t}_0 &= b_2(b_1 - b_2) - \epsilon^2 b_1(b_1 + b_2).
\end{aligned}$$

This makes $\rho\xi = \tau + u$ and $\rho\eta = \bar{t}_1\tau + \bar{t}_0 u$, where

$$\bar{t}_1 = b_1(b_1 - b_2) + \epsilon^2 b_2(b_1 + b_2).$$

On the curve we have $t = \tau/(\tau + u)$ and therefore $\eta = \bar{t}\xi$, where $\bar{t} = (\bar{t}_1 - \bar{t}_0)t + \bar{t}_0$.

The upcoming lemma uses the parameterization $\xi = \psi(\bar{t})/\bar{h}(\bar{t})$ where

$$\bar{h}(\bar{t}) = h(t) \quad \text{and} \quad \psi(\bar{t}) = \frac{b_2 - b_1}{\rho}\theta(t) + \frac{\epsilon(b_1 + b_2)}{\rho}\phi(t) + h(t).$$

Since we are assuming that $h(t) > h_{\min}$ for $0 \le t \le 1$, we have $0 < h_{\min} \le \bar{h}(\bar{t}_1) \le \lambda h_{\min} \le \lambda \bar{h}(\bar{t}_0)$.

We are now in a position to verify that the rest of the assumptions of Lemma C.3 below hold when $|b_1| > |b_2|$. The bound on $\bar{h}'(\bar{t}) = h'(t)/(\bar{t}_1 - \bar{t}_0)$ comes from the fact that $h'(0) = 3(Z_1 - Z_0)$ and $h'(1) = 3(Z_3 - Z_2)$ while a value of Z_1 or Z_2 less than $(\lambda - \frac{9}{4}(\lambda - 1))h_{\min}$ would force $h(\frac{1}{3}) < h_{\min}$ or $h(\frac{2}{3}) < h_{\min}$; e.g., $h(\frac{1}{3}) = \frac{4}{9}Z_1 + \frac{8}{27}Z_0 + \frac{2}{9}Z_2 + \frac{1}{27}Z_3 \le \frac{4}{9}Z_1 + \frac{5}{9}\lambda h_{\min}$.

Since the $(\epsilon\xi, \eta)$ coordinates are rotated and scaled by a factor of

$$|\epsilon| \frac{\sqrt{(b_2 - b_1)^2 + \epsilon^2(b_1 + b_2)^2}}{|\rho|} \le \frac{\sqrt{(b_2 - b_1)^2 + (b_1 + b_2)^2}\,|\epsilon|}{|\rho|} \le \frac{\sqrt{2}\,|\epsilon|}{|\rho|},$$

the bound $\sqrt{r^2 + s^2} \le L$ becomes $\sqrt{\epsilon^2\xi^2 + \eta^2} \le \sqrt{2}\,|\epsilon|\,L/\,|\rho|$. Similarly, the bound on the magnitude of $(\epsilon\xi'(\bar{t}), \eta'(\bar{t}))$ comes from the fact that

$$(r'(t), s'(t)) = \frac{(\theta'(t), \phi'(t))}{h(t)} - \left(\frac{h'(t)}{h(t)}\right)\frac{(\theta(t), \phi(t))}{h(t)}$$

has magnitude at most $6L + \frac{27}{4}(\lambda - 1)L$ and this needs to be scaled by $\sqrt{2}\,|\epsilon|\,L/\,|\rho|$ and divided by $\frac{d\bar{t}}{dt} = \bar{t}_1 - \bar{t}_0$.

The only other assumptions are that $\bar{t}_1 > 0$ and $\bar{t}_1 - \bar{t}_0 > 0$. These are easily verified when $|b_1| > |b_2|$ by noting that $\bar{t}_1 = (b_1^2 - b_1 b_2)(1 - \epsilon^2) + \epsilon^2(b_1^2 + b_2^2)$ and $\bar{t}_1 - \bar{t}_0 = (b_1 - b_2)^2 + \epsilon^2(b_1 + b_2)^2$.

Lemma C.3 *Let*

$$\alpha_1 = \frac{(27\lambda - 3)\,|\epsilon|\,L}{2\sqrt{2}\,|\rho|}, \qquad \alpha_2 = \frac{\sqrt{2}\,|\epsilon|\,L}{|\rho|}, \qquad \alpha_3 = \frac{27}{4}(\lambda - 1),$$

and suppose that $|b_1| > |b_2|$ *and a rational cubic curve*

$$(\epsilon\xi(\bar{t}),\, \eta(\bar{t})) = \left(\frac{\epsilon\psi(\bar{t})}{\bar{h}(\bar{t})},\, \frac{\bar{t}\psi(\bar{t})}{\bar{h}(\bar{t})}\right)$$

defined on some interval $\bar{t}_0 \le \bar{t} \le \bar{t}_1$ *with* $\bar{t}_1 > 0$ *and* $\xi(\bar{t}_0) = 1$ *has endpoints at most* α_2 *units apart. If* $0 < \bar{h}(\bar{t}_1) \le \lambda \bar{h}(\bar{t}_0)$ *and*

$$\sqrt{\epsilon^2\xi'(\bar{t})^2 + \eta'(\bar{t})^2} \le \frac{\alpha_1}{\bar{t}_1 - \bar{t}_0} \quad \text{and} \quad \left|\bar{h}'(\bar{t})\right| \le \frac{\alpha_3 \bar{h}(\bar{t})}{\bar{t}_1 - \bar{t}_0}$$

at $\bar{t} = \bar{t}_0$ *and* $\bar{t} = \bar{t}_1$*, then*

$$\left(1 - \frac{\alpha_1}{\bar{t}_1}\right)\bar{t}_1 \le \left(\frac{283\lambda}{36} - \frac{15}{4} + \frac{(27\lambda^2 - 9\lambda)\,|\epsilon|\, L}{\sqrt{2}\,\bar{t}_1\,|\rho|}\right)|\bar{t}_0|\,. \tag{55}$$

Proof. The basic idea is that since $\eta = \xi\bar{t}$, the restrictions on $\bar{h}(\bar{t})$ and the fact that $\psi(\bar{t})$ is a quadratic polynomial do not allow $\xi(\bar{t})$ to behave as it would have to if $\bar{t}_1$ were much greater than $|\bar{t}_0|$. Since $\xi(\bar{t}_0) = 1$ and therefore $|\bar{t}_1\xi(\bar{t}_1) - \bar{t}_0| = |\eta(\bar{t}_1) - \eta(\bar{t}_0)| \le \alpha_2$, we immediately have

$$|\xi(\bar{t}_1)| \le \frac{|\bar{t}_0| + \alpha_2}{\bar{t}_1}.$$

Differentiating $\eta(\bar{t}) = \bar{t}\xi(\bar{t})$ yields

$$|\xi(\bar{t}) + \bar{t}\xi'(\bar{t})| = |\eta'(\bar{t})| \le \sqrt{\epsilon^2\xi'(\bar{t})^2 + \eta'(\bar{t})^2} \le \frac{\alpha_1}{\bar{t}_1 - \bar{t}_0},$$

and solving for $\xi'(\bar{t})$ yields $|\xi'(\bar{t}_1)| \le |\xi(\bar{t}_1)|/\bar{t}_1 + \alpha_1/(\bar{t}_1^2 - \bar{t}_0\bar{t}_1)$ when $\bar{t} = \bar{t}_1$ and $|\xi'(\bar{t}_0)| \ge 1/|\bar{t}_0| - \alpha_1/\left|\bar{t}_0\bar{t}_1 - \bar{t}_0^2\right|$ when $\bar{t} = \bar{t}_0$. Note that we can assume $\bar{t}_0 \ne 0$ because otherwise $1 = |\xi(\bar{t}_0) + \bar{t}_0\xi'(\bar{t}_0)| \le \alpha_1/\bar{t}_1$ and the lemma holds because the left-hand side of (55) is nonpositive.

Since $\psi'(\bar{t}) = \bar{h}(\bar{t})\xi'(\bar{t}) + \xi(\bar{t})\bar{h}'(\bar{t})$, we have

$$\begin{aligned}
|\psi'(\bar{t}_1)| &\le \bar{h}(\bar{t}_1)\left(\left(\frac{|\xi(\bar{t}_1)|}{\bar{t}_1} + \frac{\alpha_1}{\bar{t}_1^2 - \bar{t}_0\bar{t}_1}\right) + \frac{\alpha_3\,|\xi(\bar{t}_1)|}{\bar{t}_1 - \bar{t}_0}\right)\\
&= \frac{\bar{h}(\bar{t}_1)}{\bar{t}_1 - \bar{t}_0}\left(\frac{\alpha_1}{\bar{t}_1} + \left(1 - \frac{\bar{t}_0}{\bar{t}_1} + \alpha_3\right)|\xi(\bar{t}_1)|\right)\\
|\psi'(\bar{t}_0)| &\ge \bar{h}(\bar{t}_0)\left(\left(\frac{1}{|\bar{t}_0|} - \frac{\alpha_1}{|\bar{t}_0\bar{t}_1 - \bar{t}_0^2|}\right) - \frac{\alpha_3\,|\xi(\bar{t}_0)|}{\bar{t}_1 - \bar{t}_0}\right)\\
&= \frac{\bar{h}(\bar{t}_0)}{\bar{t}_1 - \bar{t}_0}\left(\frac{\bar{t}_1 - \bar{t}_0}{|\bar{t}_0|} - \frac{\alpha_1}{|\bar{t}_0|} - \alpha_3\right).
\end{aligned}$$

Since $\psi(\bar{t})$ is a quadratic polynomial, we have

$$\psi(\bar{t}_1) - \frac{\bar{t}_1 - \bar{t}_0}{2}\psi'(\bar{t}_1) = \psi(\bar{t}_0) + \frac{\bar{t}_1 - \bar{t}_0}{2}\psi'(\bar{t}_0),$$

hence it suffices to show that we cannot have $\psi(\bar{t}_0) = \bar{h}(\bar{t}_0)$ and $|\psi(\bar{t}_1)| = \left|\xi(\bar{t}_1)\bar{h}(\bar{t}_1)\right|$ when $\bar{t}_1/|\bar{t}_0|$ is too large. From

$$\left|\psi(\bar{t}_0) + \frac{\bar{t}_1 - \bar{t}_0}{2}\psi'(\bar{t}_0)\right| \ge \frac{\bar{h}(\bar{t}_0)}{2}\left(\frac{\bar{t}_1 - \bar{t}_0}{|\bar{t}_0|} - \frac{\alpha_1}{|\bar{t}_0|} - \alpha_3 - 2\right) \ge \frac{\bar{h}(\bar{t}_0)}{2}\left(\frac{\bar{t}_1}{|\bar{t}_0|}\left(1 - \frac{\alpha_1}{\bar{t}_1}\right) - \alpha_3 - 3\right)$$

and

$$\begin{aligned}
\left|\psi(\bar{t}_1) - \frac{\bar{t}_1 - \bar{t}_0}{2}\psi'(\bar{t}_1)\right| &\le \frac{\bar{h}(\bar{t}_1)}{2}\left(\frac{\alpha_1}{\bar{t}_1} + \left(3 - \frac{\bar{t}_0}{\bar{t}_1} + \alpha_3\right)|\xi(\bar{t}_1)|\right)\\
&\le \lambda\frac{\bar{h}(\bar{t}_0)}{2}\left(\frac{\alpha_1}{\bar{t}_1} + \left(3 + \frac{|\bar{t}_0|}{\bar{t}_1} + \alpha_3\right)\frac{|\bar{t}_0| + \alpha_2}{\bar{t}_1}\right),
\end{aligned}$$

it immediately follows that

$$\frac{\bar t_1}{|\bar t_0|}\left(1-\frac{\alpha_1}{\bar t_1}\right)-\alpha_3-3\le\frac{\lambda\alpha_1}{\bar t_1}+\lambda\left(3+\frac{|\bar t_0|}{\bar t_1}+\alpha_3\right)\frac{|\bar t_0|+\alpha_2}{\bar t_1}$$

and therefore

$$\frac{\bar t_1}{|\bar t_0|}\left(1-\frac{\alpha_1}{\bar t_1}\right)\le\alpha_3+3+\frac{\lambda(\alpha_1+(\alpha_3+3)\alpha_2)}{\bar t_1}+\lambda\left(\alpha_3+3+\frac{\alpha_2}{\bar t_1}\right)\frac{|\bar t_0|}{\bar t_1}+\frac{\lambda\bar t_0^2}{\bar t_1^2}. \tag{56}$$

When $\bar t_1/|\bar t_0|\le\alpha_3+3+\alpha_2/\bar t_1$, we have

$$\left(1-\frac{\alpha_1}{\bar t_1}\right)\frac{\bar t_1}{|\bar t_0|}\le\frac{\bar t_1}{|\bar t_0|}\le\alpha_3+3+\frac{\alpha_2}{\bar t_1}=\frac{27\lambda}{4}-\frac{15}{4}+\frac{\sqrt{2}\,|\epsilon|\,L}{\bar t_1\,|\rho|}\le\frac{283\lambda}{36}-\frac{15}{4}+\frac{(27\lambda^2-9\lambda)\,|\epsilon|\,L}{\sqrt{2}\,\bar t_1\,|\rho|}$$

as required. Otherwise the last two terms in the right-hand side of (56) have upper bounds of λ and $\lambda/9$ respectively. Substituting the upper bounds for the last two terms of (56) and using the given expressions for α_1, α_2, and α_3 yields the desired result. □

Lemma C.4 *If $|\epsilon|\,L\le 0.031\,|\rho|/(\lambda-\frac{1}{9})$ then all points on the curve $\big(r(t),s(t)\big)$ satisfy*

$$|\tau+u|\ge\frac{0.031\,|\rho|}{\lambda-0.43},$$

where $\tau=rb_2+\epsilon sb_1$ and $u=-rb_1+\epsilon sb_2+\rho$.

Proof. First consider the case when $|b_2|>0.1$. Lemma C.2 implies that

$$|\tau+u|\ge\frac{|b_2\rho+\epsilon s|}{\max(|b_1|\,,|b_2|)}\ge|b_2\rho+\epsilon s|\ge|b_2\rho|-|\epsilon|\,L=\left(|b_2|-\frac{|\epsilon|\,L}{|\rho|}\right)|\rho|\,, \tag{57}$$

hence

$$|\tau+u|\ge\left(0.1-\frac{0.031}{\lambda-\frac{1}{9}}\right)|\rho|\ge\left(0.1-\frac{0.031}{\frac{8}{9}}\right)|\rho|\ge\frac{0.031}{0.57}\,|\rho|\ge\frac{0.031}{\lambda-0.43}\,|\rho|\,.$$

When $|b_2|\le 0.1$, the strategy is to use Lemma C.3 to get a lower bound on $|b_2|$ and then plug this into (57). Since

$$\bar t_1=(b_1^2-b_1b_2)(1-\epsilon^2)+\epsilon^2(b_1^2+b_2^2)$$

is a weighted average of $b_1^2-b_1b_2=0.99-b_1b_2>0.89$ and $b_1^2+b_2^2=1$, it follows that $\bar t_1>0.89$. Using $|b_1-b_2|<1.1$ in

$$\bar t_0=b_2(b_1-b_2)(1-\epsilon^2)-\epsilon^2(b_1^2+b_2^2)$$

shows that $|\bar t_0|<1.1\,|b_2|+\epsilon^2$.

The bounds on $\bar t_1$ and $\bar t_0$ make it possible to apply Lemma C.3. From $\bar t_1>0.89$ we have

$$\frac{\alpha_1}{\bar t_1}<\frac{(27\lambda-3)\,|\epsilon|\,L}{0.89\cdot 2\sqrt{2}\,|\rho|}<\frac{27\cdot 0.031}{0.89\cdot 2\sqrt{2}}<\frac{1}{3}$$

and

$$\frac{(27\lambda^2-9\lambda)\,|\epsilon|\,L}{\sqrt{2}\,\bar t_1\,|\rho|}<\frac{(27\lambda^2-9\lambda)\,|\epsilon|\,L}{0.89\sqrt{2}\,|\rho|}<\frac{27\cdot 0.031\lambda}{0.89\sqrt{2}}<\frac{2\lambda}{3},$$

hence (55) becomes

$$\begin{aligned}\frac{2}{3}0.89<\left(1-\frac{\alpha_1}{\bar t_1}\right)\bar t_1&\le\left(\frac{283\lambda}{36}-\frac{15}{4}+\frac{(27\lambda^2-9\lambda)\,|\epsilon|\,L}{\sqrt{2}\,\bar t_1\,|\rho|}\right)|\bar t_0|\\&<\left(\frac{307\lambda}{36}-\frac{15}{4}\right)|\bar t_0|<\left(\frac{307\lambda}{36}-\frac{15}{4}\right)(1.1\,|b_2|+\epsilon^2).\end{aligned}$$

Using

$$\left(\frac{307\lambda}{36} - \frac{15}{4}\right)\epsilon^2 < 0.031^2 \frac{\frac{307}{36}\lambda - \frac{15}{4}}{(\lambda - \frac{1}{9})^2} < 0.031^2 \frac{9\lambda - 1}{(\lambda - \frac{1}{9})^2} = \frac{9 \cdot 0.031^2}{\lambda - \frac{1}{9}} < 0.01$$

it follows that

$$|b_2| > \frac{\frac{2}{3}0.89 - 0.01}{1.1(\frac{307}{36}\lambda - \frac{15}{4})} > \frac{0.062}{\lambda - 0.43}.$$

Plugging this into (57) completes the proof. □

We now have a lower bound on the magnitude of $\tau + u$, but we need a lower bound on the magnitude of $(\frac{dr}{dt}, \frac{ds}{dt})$ in order to get for Lemma C.1 to give a lower bound on the gradient magnitude.

Lemma C.5 *If* $|\epsilon| L \le 0.031 |\rho| / (\lambda - \frac{1}{9})$ *then*

$$\sqrt{\left(\frac{dr}{dt}\right)^2 + \left(\frac{ds}{dt}\right)^2} \ge \frac{0.031 |\rho|}{\lambda - 0.43}$$

whenever $0 \le t \le 1$.

Proof. Since $t = \tau/(\tau + u)$ when $\tau = rb_2 + \epsilon s b_1$ and $u = -rb_1 + \epsilon s b_2 + \rho$, we have

$$\begin{aligned}\tau + u = (\tau + u)\frac{d}{dt}\left(\frac{\tau}{\tau + u}\right) &= \frac{d\tau}{dt} - \left(\frac{d\tau}{dt} + \frac{du}{dt}\right)\frac{\tau}{\tau + u} \\ &= \left(b_2 - \frac{(b_2 - b_1)\tau}{\tau + u}\right)\frac{dr}{dt} + \epsilon\left(b_1 - \frac{(b_1 + b_2)\tau}{\tau + u}\right)\frac{ds}{dt}.\end{aligned}$$

Thus $\tau + u$ is the dot product of $(\frac{dr}{dt}, \frac{ds}{dt})$ and a vector

$$\big(b_2 - (b_2 - b_1)t,\ \epsilon(b_1 - (b_1 + b_2)t)\big)$$

which lies on the line segment between $(b_2, \epsilon b_1)$ and $(b_1, -\epsilon b_2)$ and therefore has length at most one. This forces the length of $(\frac{dr}{dt}, \frac{ds}{dt})$ to be at least $|\tau + u|$, and Lemma C.4 gives the desired lower bound. □

Notice that Lemma C.5 gives a lower bound on the order of $|\rho|$, while the obvious upper bound on the magnitude of $(\frac{dr}{dt}, \frac{ds}{dt})$ is a constant times L. The disparity reflects the fact that when ρ/L and ϵ are both small, the curve can have "flat spots" such as those that lead to relatively large errors in Section 3.1. In spite of the extra factor of $|\rho|/L$, Lemmas C.1, C.4, and C.5 give a good bound on the gradient:

Theorem C.6 *If* $|\epsilon| L \le 0.031 |\rho| / (\lambda - \frac{1}{9})$ *then the function* $G(r, s)$ *defined in Section 2.2 has*

$$\sqrt{\left(\frac{dG}{dr}\right)^2 + \left(\frac{dG}{ds}\right)^2} \ge \frac{0.031^2 |\rho|^3 h_{\min}}{(\lambda - 0.43)^2}$$

on the curve $\big(r(t), s(t)\big)$ *for* $0 \le t \le 1$.

References

[1] B. W. Char, K. O. Geddes, G. H. Gonnet, M. B. Monagan, and S. M. Watt. *MAPLE Reference Manual.* WATCOM, Waterloo, Ontario, 1988.

[2] R. T. Farouki and V. T. Rajan. On the numerical condition of polynomials in Bernstein form. *Computer-Aided Geometric Design*, 4:191–216, 1987.

[3] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, Maryland, 1983.

[4] John D. Hobby. Rasterization of nonparametric curves. *ACM Transactions on Graphics*, 9(3):262–277, July 1990.

[5] Richard R. Patterson. Projective transformation of the parameter of a Bernstein-Bézier curve. *ACM Transactions on Graphics*, 4(4):276–300, 1985.

[6] T. W. Sederberg. *Implicit and Parametric Curves and Surfaces for Computer Aided Geometric Design*. PhD thesis, Dept. of Mechanical Engineering, Purdue University, 1983.

[7] T. W. Sederberg. Planar piecwise algebraic curves. *Computer-Aided Geometric Design*, 1:241–255, 1984.

[8] T. W. Sederberg and D. C. Anderson. Implicit representation of parametric curves and surfaces. *Computer Vision Graphics and Image Processing*, 28(1):72–84, 1984.

[9] T. W. Sederberg, D. C. Anderson, and R. N. Goldman. Implicitization, inversion, and intersection of planar rational cubic curves. *Computer Vision Graphics and Image Processing*, 31(1):89–102, July 1985.

[10] T. W. Sederberg and R. N. Goldman. Algebraic geometry for computer-aided geometric design. *IEEE Computer Graphics and Applications*, 6(6):52–59, 1986.

[11] T. W. Sederberg and R. N. Goldman. Analytic approach to intersection of all piecewise parametric rational cubic curves. *Computer-Aided Design*, 19(6):282–292, 1987.

[12] T. W. Sederberg and S. R. Parry. Comparison of three curve intersection algorithms. *Computer-Aided Design*, 18(1):58–63, 1986.

[13] J. G. Semple and L. Roth. *Algebraic Geometry*. Oxford, 1949.