Chapter 5

Representation of Curved Edges and Faces

The focus of this chapter is the question of how to represent curved surface elements in boundary representation. Simple conventions suffice for the planar situation. For example, the geometric locus of an edge can be specified by the coordinates of its two bounding vertices. Curved boundary elements are not that simple, and we have to give a two-part description, consisting of a surface or space curve, the *carrier* of the face or edge, and a *boundary description* that delimits a subarea or segment. In the case of planar faces, the carrier is the plane containing the face and the boundary description is an edge graph. The carrier for an edge of a polyhedron is a line and can be inferred from the vertices.

In this chapter, we explore some of the issues arising when specific conventions are chosen for the two parts. In the case of the carrier specification, we discuss ways in which the surface can be given, and explain techniques for converting between them. When specifying the bounding structure of edges and faces, some geometric problems must be considered to avoid ambiguities. These are also discussed.

We restrict attention to algebraic surfaces and curves. This restriction is reasonable in the sense that the class is very rich and includes most of the curves and surfaces used in, for instance, engineering design. In particular, this class includes all the major parametric surfaces used in geometric modeling, including Bezier surfaces, nonuniform rational B-splines, and so on. It does not include all surfaces and space curves, however. For example, the helix is not an algebraic space curve.

5.1 Chapter Overview

We begin with a brief review of affine versus projective space. Although one ordinarily considers geometric objects in affine spaces only, there is a technical advantage to considering them also in projective space. For example, the question whether a curve has a rational parametric form cannot be decided unless curve properties “at infinity” are taken into account, and these properties are revealed when the curve is considered in projective space.

We then explain basic properties of implicit and parametric curve and surface representations. Both methods have distinct and complementary advantages. In the case of implicit surfaces, it is straightforward to decide whether a given point in space is or is not on the surface. For a parametric surface, on the other hand, it is easy to generate points that lie on the surface.

Because of such complementary strengths, the problem of how to convert from one form to the other is of great practical interest. General techniques exist for converting from parametric to implicit form, at least in principle, and we review here a simple version based on the Sylvester resultant. In Chapter 7, we show how to use Gröbner bases techniques for this purpose.

Whereas the conversion from parametric to implicit form is always possible, the conversion from an implicit to a parametric form depends on specific properties not shared by all algebraic curves and surfaces. These properties are fairly technical in nature and determining them algorithmically is difficult, so we omit this characterization. Instead, we give several methods for parameterization that are applicable to restricted classes of curves and surfaces.

Parameterization of quadratic curves and surfaces is a classical problem, and we give two different methods. We then discuss in detail the parameterization of cubic curves. Some higher-degree curves and surfaces are easy to parameterize, including the class of *monoids*. For this reason, monoids have been proposed by some authors as a basic shape element in geometric modeling. We discuss monoid parameterization also.

Up to this point, the material deals with representations of the carrier of edges and faces. The identification of edges and faces on curved carriers raises problems not encountered in the polyhedral domain, as discussed toward the end of the chapter. The problem here is that the geometric and topological structure of the carrier creates opportunities for ambiguities in that, for example, the specification of an edge as a segment bounded by two points on a space curve could be interpretable in different ways. In consequence, additional data are needed to disambiguate the representation.

5.2 Affine and Projective Spaces

We will use both affine and projective spaces. *Affine* n-dimensional space is the familiar n-space.[1] Using Cartesian coordinates, a point in this space has coordinates

$$(x_1, x_2, ..., x_n)$$

where the x_i are always finite. Ordinarily, the coordinate values of the x_i are real numbers, but for some results from algebraic geometry we must also consider points with complex coordinates. When this fact is critical, we will mention it explicitly.

Projective n-dimensional space consists of points with $n+1$ coordinates $(x_0, x_1, x_2, ..., x_n)$, where not all x_k are zero. Again, each coordinate value is finite. Moreover, for nonzero numbers t, both $(x_0, x_1, ..., x_n)$ and $(tx_0, tx_1, ..., tx_n)$ describe the same point. The variable x_0 is sometimes called the *homogenizing* variable. We will usually write it as the first coordinate. However, any one of the other variables could be considered to be the homogenizing variable, a fact we will illustrate further.

As before, we may have to consider complex coordinates. In projective space, the points $(0, x_1, ..., x_n)$ are said to be *points at infinity*. These points form the hyperplane $x_0 = 0$. In particular, for $n = 2$, the points at infinity comprise the *line* at infinity.

Affine n-space can be considered a restriction of projective n-space by requiring $x_0 \neq 0$. In this sense, we might say that affine space is the finite part of projective space. In turn, projective n-space can be embedded into affine $n+1$ space as follows: Consider each point $(a_0, a_1, a_2, ..., a_n)$ of projective space as the line

$$\begin{aligned} x_1 &= a_1 t \\ x_2 &= a_2 t \\ &\vdots \\ x_n &= a_n t \\ x_{n+1} &= a_0 t \end{aligned}$$

In consequence, projective n-space is the space of all lines in affine $(n+1)$-space that contain the origin. The restriction of projective n-space to affine n-space may now be considered to be all points in which the embedded line space intersects the plane $x_{n+1} = 1$. Figure 5.1 illustrates the embedding of two-dimensional projective space into three-dimensional affine space. In the figure, the point $P = (a_0, a_1, a_2)$ of projective 2-space corresponds to the line l through the origin of affine 3-space. The point $p = (a_1/a_0, a_2/a_0)$

[1] Euclidian space, considered in Chapter 2, Section 2.4, is affine space endowed with the Euclidian distance metric.

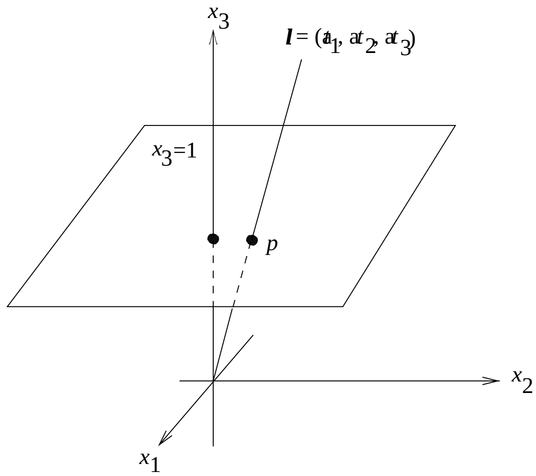

Figure 5.1 Embedding Projective Space into Affine Space

corresponding to P in affine 2-space is embedded into affine 3-space as the intersection of the line l with the plane $x_3 = 1$.

In affine spaces, the *origin* $(0, 0, ..., 0)$ is a distinguished point. In projective n-space, we distinguish $n + 1$ *fundamental* points with the coordinates

$$\begin{array}{c} (1, 0, ..., 0, 0) \\ (0, 1, ..., 0, 0) \\ \vdots \\ (0, 0, ..., 1, 0) \\ (0, 0, ..., 0, 1) \end{array}$$

For example, in the projective plane, the point $(1, 0, 0)$ is the affine origin, the point $(0, 1, 0)$ is the intersection of the x axis with the line at infinity, and the point $(0, 0, 1)$ is the intersection of the y axis with the line at infinity. The fundamental points span the *tetrahedron of reference.*

An *affine transformation* is a linear transformation of the form

$$\begin{array}{rcl} y_1 & = & a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n + b_1 \\ y_2 & = & a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n + b_2 \\ & \vdots & \\ y_n & = & a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n + b_n \end{array}$$

Intuitively, an affine transformation may shear or stretch a geometric shape.

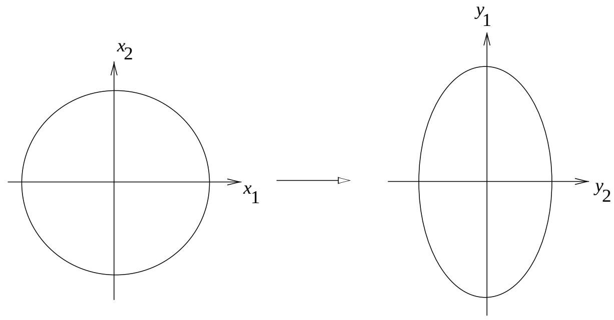

Figure 5.2 Affine Transformation of a Circle

Simple examples include rotations and reflections. Note that the matrix

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

can be singular, in which case the transformation achieves a parallel projection.

Example 5.1: The affine transformation

$$\begin{aligned} y_1 &= x_1 \\ y_2 &= x_1 + x_2 \end{aligned}$$

changes the circle $x_1^2 + x_2^2 - 1$ into the ellipse $2y_1^2 - 2y_1y_2 + y_2^2 - 1$. See also Figure 5.2. $\diamond$

A *projective transformation* is a linear transformation of the form

$$\begin{aligned} y_0 &= a_{00}x_0 + a_{01}x_1 + a_{02}x_2 + ... + a_{0n}x_n \\ y_1 &= a_{10}x_0 + a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \\ y_2 &= a_{20}x_0 + a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \\ &\vdots \\ y_n &= a_{n0}x_0 + a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n \end{aligned}$$

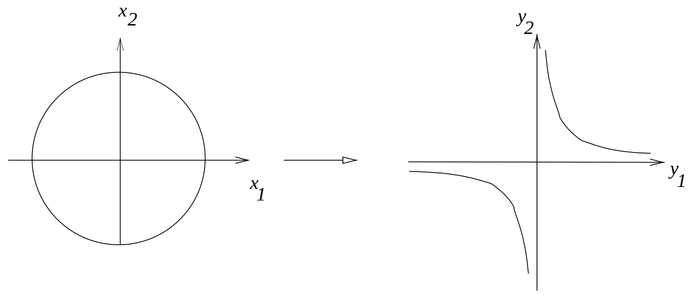

Figure 5.3 Projective Transformation of a Circle

It is well known that a projective transformation can change a circle into any conic section. Singular projective transformations are routinely used in computer graphics for computing perspective images.

Example 5.2: The projective transformation

$$\begin{aligned} y_0 &= x_2 \\ y_1 &= (x_0 + x_1)/2 \\ y_2 &= (x_0 - x_1)/2 \end{aligned}$$

changes the circle $x_1^2 + x_2^2 - x_0^2$ into the hyperbola $y_0^2 - 4y_1y_2$, shown in Figure 5.3. $\diamond$

5.3 Implicit Representations

5.3.1 Implicit Surfaces

Every algebraic surface in affine 3-space is determined by an implicit equation

$$f(x, y, z) = 0$$

where $f(x, y, z)$ is a polynomial in the unknowns x, y, and z. The surface consists of all points (x, y, z) that satisfy this equation. In solid modeling, real coordinates are considered. However, to apply results from algebraic geometry, we must allow complex coordinates; see also Section 7.2.1 in Chapter 7.

The surface is *irreducible* if f does not factor over the field of complex numbers; that is, if there do not exist two nonconstant polynomials $h(x, y, z)$ and $k(x, y, z)$, possibly with complex coefficients, such that

$f(x, y, z) = h(x, y, z)k(x, y, z)$. A surface that is not irreducible is *reducible*. For example, the cylinder $x^2 + y^2 - 1 = 0$ is an irreducible surface, but $x^2 + y^2 = 0$ is a pair of planes, for $x^2 + y^2 = (x + iy)(x - iy)$. The planes have points with complex coordinates and intersect in the real line $x = y = 0$. Note that, over the field of real numbers, $x^2 + y^2$ does not factor. Determining algorithmically whether a given polynomial $f(x, y, z)$ factors over the field of complex numbers is difficult.

Reducibility of a surface $f = 0$ means geometrically that the surface can be decomposed into two separate surfaces, each of which can be described separately by an implicit equation. This requires examining the surface in complex space, since we can find examples of surfaces in real three-dimensional space that appear to consist of two disjoint components that are, in fact, connected when complex surface points are considered.

The *gradient* or *normal* vector of the surface $f(x, y, z) = 0$ is the vector (f_x, f_y, f_z), where f_x, f_y, and f_z are the partial derivatives of f by x, y, and z, respectively. For example, the gradient of the sphere $f = x^2 + y^2 + z^2 - 1 = 0$ at the point (x, y, z) is $(2x, 2y, 2z)$. So, at $(1, 0, 0)$, the gradient is $(2, 0, 0)$. Sometimes, the gradient vector is *normed* to length 1. A point (x_0, y_0, z_0) on an irreducible surface $f = 0$ is *regular* if the gradient at the point is not the zero vector. Otherwise, the point is *singular*.

For every surface point, there exists a *tangent space* to the surface, consisting of all tangent lines to the surface at that point. It can be proved that, at a regular point, the tangent space is a plane, called the *tangent plane*. At a singular point, the tangent space is a *cone*; that is, it is a surface generated by lines each containing the singular point.

Assume that the surface $f = 0$ contains the origin. It is not difficult to show that, at the origin, the equation of the tangent space is given by the terms of lowest degree in f. For example, the sphere $x^2 + y^2 + z^2 - 2x = 0$ contains the origin and has at that point the tangent plane $2x = 0$.

The terms of lowest order are called the *initial form* of f. If the origin is a regular point of $f = 0$, then the initial form is linear; that is, the tangent space equation is that of a plane. If the origin is singular, then the initial form of f is nonlinear and describes the tangent cone.

Consider the initial form $h(x, y, z)$ of $f(x, y, z)$, and assume that the origin is a singular surface point. All terms in h have equal degree $d > 1$. Consider any point $p = (a, b, c)$ on the surface $h = 0$, where p is not the origin. Then the point $q = (ta, tb, tc)$ is also on $h = 0$ for all values of t, since $h(q) = t^d h(p)$. It follows that h contains the line

$$\begin{aligned} x &= ta \\ y &= tb \\ z &= tc \end{aligned}$$

Note that this line contains the origin. Since the line is constructed with an arbitrary point $p \neq (0,0,0)$ on $h = 0$, we have shown that the initial form defines a cone whose vertex is the origin.

The initial form gives information about the surface geometry at the origin. Similarly, the *degree form* of *f*, consisting of all terms of highest degree in *f*, yields information about the surface behavior at infinity.

The polynomial f describing a surface $f = 0$ is not a unique description of the surface, since $cf(x,y,z) = 0$ also describes the surface, provided that c is not zero. For this reason, a surface of degree n can be considered as a point in projective m-space, where $m + 1$ is the number of possible coefficients; that is,

$$m = \binom{n+3}{3} - 1 = \frac{n(n^2 + 6n + 11)}{6}$$

For example, to each quadric in 3-space there corresponds a point in projective 9-space, since a quadric is specified by the ratio of 10 coefficients.

The formula for m is derived as follows. Let $T(n,k)$ be the number of terms of degree up to n that can be formed with k variables. Clearly, $T(n,1) = n+1$. When forming all terms with $k+1$ variables, we can group them by x^j, where x is one of the variables. Then the group for x^j consists of all terms $x^j u$, where u is formed with k variables and has degree $0, 1, \ldots, n-j$. The possible terms u are, therefore, all terms of degree up to $n - j$ that are formed with k variables, so

$$T(n, k+1) = \sum_{j=0}^{n} T(j,k)$$

By induction, one shows easily that

$$T(n,k) = \binom{n+k}{k}$$

Note that $m = T(n,3) - 1$.

If f is multiplied with a polynomial g, then the zeros of the product $g(x,y,z)f(x,y,z) = 0$ are of the union of the zeros of $f = 0$ and of the zeros of $g = 0$. When g is varied, only the zeros of f are in every zero set. This motivates defining the surface $f = 0$ as the set of *common* zeros of all polynomials of the form $g(x,y,z)f(x,y,z)$, where g is any polynomial, including the trivial polynomial c, where c is a constant. The set of all such polynomials is an *ideal* — more precisely, a *principal ideal* — and f is a *generator* of the ideal. Ideals will be discussed in Chapter 7.

The polynomial $f(x,y,z) = 0$ describes a surface in affine 3-space. The corresponding surface in projective 3-space is obtained through *homogenizing* the polynomial f; that is, by substituting x/w for x, y/w for y, and z/w for z in f, followed by clearing the denominators.[2] The resulting polynomial

[2] Note that we use the same coordinate variables for the affine and the projective spaces.

$F(x, y, z, w)$ is the *homogeneous form* of f. All of its terms are of equal order. Similarly, the affine form can be obtained from F by substituting 1 for w in F. For example, the affine quadric

$$x^2 + 2x + y^2 + z^2 - 1 = 0$$

is homogenized as

$$x^2 + 2xw + y^2 + z^2 - w^2 = 0$$

We have used w as the additional, *homogenizing* variable. However, by a simple homogeneous transformation, we can rename variables. In effect, the embedding of affine space into projective space is changed by such a transformation. For example, consider $x = 0$ to be the plane at infinity. Then the affine part of projective space consists of the points

$$\left\{ \left(\frac{y}{x}, \frac{z}{x}, \frac{w}{x} \right) \;\middle|\; x \neq 0 \right\}$$

so the finite part of the surface $x^2 + 2xw + y^2 + z^2 - w^2 = 0$ is the surface

$$1 + 2w + y^2 + z^2 - w^2 = 0$$

in affine (y, z, w)-space. This may change the shape of the surface since now, as it were, we "see" a different finite part of it. In our example, the sphere $x^2 + 2x + y^2 + z^2 - 1 = 0$, in (x, y, z)-space, has been changed to the hyperboloid $1 + 2w + y^2 + z^2 - w^2 = 0$, in (y, z, w)-space.

5.3.2 Implicit Curves

An algebraic space curve is the common intersection of two or more surfaces. Although solid modeling usually restricts attention to those space curves that are the intersection of just two surfaces, one should remember that certain space curves cannot be defined algebraically as the intersection of only two surfaces.[3]

As in the case of surfaces, a space curve can be understood as the set of common zeros of all polynomials of the form $u_1 f_1 + u_2 f_2 + \cdots + u_k f_k$, where the u_i are arbitrary polynomials in x, y, and z, and the f_i are fixed polynomials defining the intersecting surfaces. The polynomials of this form constitute the ideal generated by the f_i.

[3] There are subtleties in this statement that are discussed in Chapter 7, Sections 7.2.5 and 7.2.6.

A *rational map* between two projective spaces of the same dimension is a map

$$\begin{aligned} y_0 &= F_0(x_0, ..., x_n) \\ y_1 &= F_1(x_0, ..., x_n) \\ &\vdots \\ y_n &= F_n(x_0, ..., x_n) \end{aligned}$$

where the F_i are homogeneous polynomials of the same degree. It induces the rational map

$$y_i = \frac{F_i(1, x_1, ..., x_n)}{F_0(1, x_1, ..., x_n)}$$

between the embedded affine spaces. A rational map is *birational* if it is invertible; that is, if there exists an inverse rational map. Simple examples of birational maps are provided by the (rational) parametric representation of certain curves and surfaces. Here, a parametric curve is in birational correspondence with a line, and a parametric surface is in birational correspondence with the plane. From algebraic geometry, we know that every algebraic space curve is in birational correspondence with some plane algebraic curve. As we shall see, this fact plays a role in some surface-intersection algorithms.

5.3.3 Bezout's Theorem

Algebraic geometry has studied the relationship between the degree of an algebraic curve and the number of points in which that curve intersects another algebraic curve. The first theorem of this kind is due to Bezout and is as follows.

> **Theorem**
> Let f and g be two algebraic curves of degree m and n, respectively. If f and g intersect in more than mn points, then they have a common component.

In consequence, two curves that do not share a common component have at most mn intersection points. By assigning multiplicities to some of these intersections, we can put the theorem into a stronger form.

> **Theorem**
> Let f and g be two algebraic curves of degree m and n, respectively. Then f and g intersect in exactly mn points, or they have a common component.

In this form, Bezout's theorem is valid only if we consider the complex curve points, as well as curve points at infinity. In the parlance of Chapter 7, we must consider the curves in projective space over an algebraically closed ground field.

We can use Bezout's theorem to explain when a curve point is a multiple point. Take a point u on the plane algebraic curve f of degree n and consider a set of lines through u. By Bezout's theorem, most of these lines intersect f in n points. If a line intersects f in $n-1$ additional points, then u is a *simple* or a *regular* point. A point that is not simple is a *multiple* or *singular* point. Note that it is not appropriate to infer multiplicities from the graphs of the curves in real affine space.

It is known that an algebraic curve has only a finite number of points that are not simple, and that a line has no multiple points. So, if all lines through u intersect f in less than $n-1$ additional points, then u is a multiple point. This definition of a multiple point is useless, however, because of the following.

Consider turning a line l through the curve point u. As the line rotates, centered at u, it intersects f in a fixed number of additional points, say $n-m$, where m is the multiplicity of u, except for finitely many positions at which l intersects f in less than $n-m$ additional points. Each such exceptional position defines a *tangent* to f at u.[4] If u is a regular point, there is only one such exceptional position. If u is a multiple point, then there could be up to m different exceptional positions, where m is the multiplicity of the point. Therefore, we define *point multiplicity* as follows.

The point u on the curve f has *multiplicity* m if an infinite number of lines through u intersect f in $n-m$ additional points. In particular, if infinitely many lines intersect f in $n-2$ additional points, then u is a *double point*. Again, this procedure makes sense only if complex as well as real curve points are considered, and when intersections at infinity are considered.

Bezout's theorem can be generalized to surfaces and space curves as follows.

Theorem
An algebraic space curve of degree m intersects an algebraic surface of degree n in mn points unless a curve component is contained in the surface. Two algebraic surfaces of degree m and n, respectively, intersect in an algebraic curve of degree mn unless they have a common component.

As before, we must consider the curves and surfaces in complex projective space.

[4] Let f be a curve that contains the origin. Then we can show that the exceptional line positions are given by the roots of the initial form of f.

5.4 Parametric Representations

5.4.1 Parametric Surfaces

Some, but not all, algebraic surfaces possess a parametric representation. Such a representation consists of three functions:

$$\begin{aligned} x &= h_1(s,t) \\ y &= h_2(s,t) \\ z &= h_3(s,t) \end{aligned}$$

For specific values of s and t, these functions assign the coordinates of a surface point in (x, y, z)-space. For example, the unit sphere can be parameterized by

$$\begin{aligned} x &= \frac{1-s^2-t^2}{1+s^2+t^2} \\ y &= \frac{2s}{1+s^2+t^2} \\ z &= \frac{2t}{1+s^2+t^2} \end{aligned}$$

From the parameter values $s = t = 1$, we obtain, for instance, the point $(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$ on the sphere.

In the example of the sphere, the parameterization does not "reach" the point $(-1, 0, 0)$, unless infinite values of s and t are permitted. We refer to such points as *singularities* of the parameterization. Infinite parameter values raise computational problems. Later, we give a projective parameterization of the sphere that avoids such singularities.

We view a parametric representation as a map from the (s, t)-plane to the surface in (x, y, z)-space. Most of the time, this map will be *rational*; that is, the functions h_1, h_2, and h_3 will be ratios of polynomials in s and t. In special situations, they can be polynomial. A mathematical characterization of when a rationally parameterizable surface has, in fact, a polynomial parameterization is a nontrivial problem. Note, however, that a rational parameterization of a surface in affine (x, y, z)-space corresponds to a polynomial parameterization of the same surface in projective (w, x, y, z)-space. The projective unit sphere is defined parametrically as

$$\begin{aligned} x &= 1-s^2-t^2 \\ y &= 2s \\ z &= 2t \\ w &= 1+s^2+t^2 \end{aligned}$$

We typically view (s,t)-space as an affine plane, expecting that distinct pairs (s,t) correspond to distinct surface points. We call such a parameterization an *affine* parameterization. On occasion, we will want a *projective* parameterization. Then the map is between a projective plane with, say, (r,s,t) coordinates and the surface in (x,y,z)-space. For $u \neq 0$, (r,s,t) and (ur,us,ut) yield the same surface point, since both coordinate triples refer to the same point in the projective plane. A projective parameterization of the affine unit sphere is

$$\begin{aligned} x &= \frac{r^2 - s^2 - t^2}{r^2 + s^2 + t^2} \\ y &= \frac{2rs}{r^2 + s^2 + t^2} \\ z &= \frac{2rt}{r^2 + s^2 + t^2} \end{aligned}$$

A projective parameterization of the projective unit sphere is

$$\begin{aligned} x &= r^2 - s^2 - t^2 \\ y &= 2rs \\ z &= 2rt \\ w &= r^2 + s^2 + t^2 \end{aligned}$$

Note that the parameter triple $(r,s,t) = (0,1,1)$ is mapped to the point $(-1,0,0)$ on the affine sphere that could not be reached with finite (s,t) values by the affine parameterization.

5.4.2 Parametric Curves

Some, but not all, algebraic curves possess a parametric representation. For example,

$$\begin{aligned} x(t) &= \frac{1 - t^2}{1 + t^2} \\ y(t) &= \frac{2t}{1 + t^2} \end{aligned}$$

is an affine parameterization of the unit circle in the affine plane. Again, we consider the parametric representation as a map, from a line with coordinate t to a curve in (x,y)-space. The point $(-1,0)$ on the circle is a singularity for this parameterization.

Again, the curve can be projectively parameterized. The projective parameterization of the circle is given by

$$\begin{aligned} x(r,t) &= \frac{r^2 - t^2}{r^2 + t^2} \\ y(r,t) &= \frac{2rt}{r^2 + t^2} \end{aligned}$$

The pair (r,t) defines a point on a projective line. The projective parameterization maps the point $(r,t) = (0,1)$ to the singular point of the affine parameterization.

As an example of a parametric representation of a space curve, we mention the twisted cubic:

$$\begin{aligned} x(t) &= t \\ y(t) &= t^2 \\ z(t) &= t^3 \end{aligned}$$

A curve or surface parameterization is *faithful* if all but finitely many distinct parameter values correspond to distinct curve or surface points. For example, the parameterization

$$\begin{aligned} x(s) &= -\frac{s^4 + 2s^3 + 3s^2 + 2s}{s^4 + 2s^3 + 3s^2 + 2s + 2} \\ y(s) &= \frac{2(s^2 + s + 1)}{s^4 + 2s^3 + 3s^2 + 2s + 2} \end{aligned} \tag{5.1}$$

of the unit circle is not faithful. To see this, observe that with $t = s^2 + s + 1$ we obtain from equation (5.1) the familiar parameterization of the circle. Thus, for all t, we have

$$(x(s_0), y(s_0)) = (x(s_1), y(s_1))$$

where s_0 and s_1 are the roots of $s^2 + s + 1 = t$.

5.4.3 Computer-Aided Geometric Design

In computer-aided geometric design (CAGD), there is a rich body of knowledge about special classes of parametric curves and surfaces. These classes are typically defined as linear combinations of certain *base functions*. Examples are Bezier curves and surfaces, and B-spline curves and surfaces. The importance of these curve and surface classes, briefly, stems from the following:

1. There is a method for constructing a curve or surface from a certain polygon or polyhedron, such that the shape of the polygon (polyhedron) gives a geometric intuition of, and control over, the shape of the curve (surface).

2. There are a number of elegant methods for evaluating and manipulating such curves and surfaces.

3. There are algorithms for aggregating larger curves or surfaces from patches of individual parametric curves or surfaces such that smoothness conditions between the patches are satisfied.

An in-depth discussion of these classes and their associated algorithms is beyond the scope of this book. At the end of this chapter, we cite a number of references on the subject.

5.5 Conversion from Implicit to Parametric Form

A useful capability in solid modeling is the conversion between implicit and parametric surface representations, since each form has different inherent strengths. Whereas all curves and surfaces with a rational parametric form can be converted to implicit form, at least in principle, not all implicit algebraic curves and surfaces possess a rational parametric form. In the case of curves, a complete characterization is given by *Noether's theorem.*

Theorem
A plane algebraic curve $f(x, y) = 0$ possesses a rational parametric form iff f has genus 0.

Roughly speaking, the curve genus measures the difference between the actual number of double points of f and the maximum number of double points a curve of the same degree as f may have. One knows that a plane curve of degree n can have no more than $(n-1)(n-2)/2$ double points, and this fact has an elementary proof. However, *counting* the number of double points of f is more subtle and involves the behavior at infinity as well as the internal structure of singular points. Algorithms for determining the genus exist but are nontrivial.

A similar characterization exists for surfaces, first given by Castelnuovo. For this characterization, two surface invariants are defined from which necessary and sufficient conditions for the existence of a parametric form are formulated. The invariants are not easily portrayed in intuitive terms.

The proof of Noether's theorem does not provide an efficient or simple computation for deriving a curve parameterization. However, for curves of

	Implicit Form	Parametric Form	
Circle	$x^2 + y^2 - r^2 = 0$	$x(t) = r\,\frac{1-t^2}{1+t^2}$	$y(t) = r\,\frac{2t}{1+t^2}$
Ellipse	$\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0$	$x(t) = a\,\frac{1-t^2}{1+t^2}$	$y(t) = b\,\frac{2t}{1+t^2}$
Hyperbola	$\frac{x^2}{a^2} - \frac{y^2}{b^2} - 1 = 0$	$x(t) = a\,\frac{1+t^2}{1-t^2}$	$y(t) = b\,\frac{2t}{1-t^2}$
Parabola	$y^2 - 2px = 0$	$x(t) = \frac{t^2}{2p}$	$y(t) = t$

Table 5.1 Standard Parameterization of Conics

degree 2 and 3, and for curves of special types such as *monoids*, simple techniques do exist. We define monoids later and describe how to parameterize them. Little appears to be known about the numerical behavior of these techniques, and the literature on this subject customarily assumes exact arithmetic.

5.5.1 Conics

We can use two basic approaches when parameterizing a conic:

1. Transform the conic into one for which a parametric form is already known, and then transform back this standard parameterization.

2. Parameterize the curve by a pencil of lines (defined later) through some curve point.

The first method requires a coordinate transformation of the conic that uses standard methods from linear algebra. The second method, with an appealingly simple underlying geometric idea, uses a modicum of coordinate transformations to simplify the implicit equation of the conic. It requires knowledge of a real curve point.

Since conics are well understood, the parameterizations for circle, ellipse, hyperbola, and parabola are well known when the curves are positioned suitably. We list them in Table 5.1.

A simple strategy for parameterizing a given conic is, therefore, first to transform the coordinate system so that the conic is properly positioned, then to retrieve a standard parameterization, and finally to apply the inverse transformation to the parametric representation.

First Method of Conic Parameterization

Any nondegenerate conic can be transformed into one of the conics in Table 5.1, using translations and rotations of the coordinate system. Formulae for computing the necessary transformation directly from the coefficients of the implicit curve equation are known, but are not reproduced here. Instead, we develop a more general method based on projective transformations, since it generalizes directly to quadric surfaces. Note, however, that projective transformations do not necessarily preserve the type of the conic. For example, it is possible that an ellipse is mapped to a hyperbola. For the purpose of curve parameterization, this is immaterial.

The general implicit conic equation, in homogeneous form, is

$$a_{11}x^2 + a_{22}y^2 + a_{33}w^2 + 2a_{12}xy + 2a_{13}xw + 2a_{23}yw = 0$$

It can be written as the *bilinear form*

$$(x\ y\ w)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}\begin{pmatrix} x \\ y \\ w \end{pmatrix} = 0$$

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

be the coefficient matrix of the conic. We seek a nonsingular matrix T such that $B = T^{-1}AT$ is diagonal. If the matrix A does not have full rank, then some diagonal elements of B will be zero. Since A is symmetric, it can be shown that such a matrix T exists and is real-valued.

The matrix T is a coordinate transformation, mapping the point $(x\ y\ w)$ to the new point $(x_1\ y_1\ w_1) = (x\ y\ w)T$. A conceptually simple method for finding it is to apply separate *Jacobi rotations* R, each designed to zero an off-diagonal element. For example, the element a_{12} is canceled by a rotation about the w axis of the form

$$R = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The element a_{13} is canceled by a rotation about the y axis, and the element a_{23} by a rotation about the x axis. The rotation matrices can be found as follows. Let

$$A' = \begin{pmatrix} m & p \\ p & n \end{pmatrix}$$

be the 2×2 submatrix containing the element p we wish to cancel. Apply the rotation matrix

$$R^T = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}$$

where $c = \cos(\alpha)$ and $s = \sin(\alpha)$. Then the $(1,2)$-element becomes

$$(R^T A' R)_{1,2} = (R^T A' R)_{2,1} = p(c^2 - s^2) - cs(n - m) = 0$$

Hence,

$$\frac{2cs}{c^2 - s^2} = \frac{2p}{n - m}$$

That is, to zero the element $(R^T A' R)_{1,2}$ in the matrix, we must choose an angle α such that

$$\tan(2\alpha) = \frac{2p}{n - m}$$

If $m = n$, then $\alpha = 45°$. Note that the angle α can always be restricted to be between $-45°$ and $45°$.

Let

$$B = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

be the final diagonal matrix obtained. We distinguish the following cases:

- Rank 3; $\lambda_1, \lambda_2, \lambda_3 \neq 0$. The conic to be parameterized is irreducible.
- Rank 2; $\lambda_i, \lambda_j \neq 0$, $\lambda_k = 0$. The conic consists of two distinct lines.
- Rank 1; $\lambda_i \neq 0$, $\lambda_j, \lambda_k = 0$. The conic consists of two coincident lines.

If the conic consists of lines, then the original conic is reducible. In this case, each component should be parameterized separately as a line; only the rank 3 case is of interest.

The standard parameterization for the nondegenerate case depends on the signs of the λ_i. If all λ_i have the same sign, then the conic is imaginary. It is not possible to transform an imaginary conic to a real-valued one, or vice versa, since we apply real-valued rotation matrices. Hence, the original conic is also imaginary, so this case is not of interest.

If only λ_1 and λ_2 have the same sign, then the transformed conic is an ellipse or a circle. With $\mu_i = 1/\sqrt{|\lambda_i|}$, the conic is parameterized by

$$\begin{aligned} x(t) &= (1 - t^2)\mu_1 \\ y(t) &= 2t\mu_2 \\ w(t) &= (1 + t^2)\mu_3 \end{aligned}$$

If λ_1 has the opposite sign of λ_2 and λ_3, then the transformed conic is a hyperbola. With $\mu_i = 1/\sqrt{|\lambda_i|}$, it is parameterized by

$$\begin{aligned} x(t) &= (1+t^2)\mu_1 \\ y(t) &= 2t\mu_2 \\ w(t) &= (1-t^2)\mu_3 \end{aligned}$$

Example 5.3: Consider the hyperbola $x^2+4xy+3y^2-4=0$. Its matrix is

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & -4 \end{pmatrix}$$

Only a_{12} needs to be canceled. We determine

$$\alpha = \frac{1}{2}\arctan(2) \approx 31.172^\circ$$

which yields as rotation matrix

$$R^T = \begin{pmatrix} 0.851 & -0.526 & 0 \\ 0.526 & 0.851 & 0 \\ 0 & 0 & -4 \end{pmatrix}$$

Then $B = R^T AR$ is

$$B = \begin{pmatrix} -0.236 & 0.0 & 0 \\ 0.0 & 4.236 & 0 \\ 0 & 0 & -4 \end{pmatrix}$$

It is diagonal and represents a hyperbola. As an affine curve, this conic is parameterized by

$$\begin{aligned} x_1(t) &= 4.117\,\frac{1+t^2}{1-t^2} \\ y_1(t) &= 0.972\,\frac{2t}{1-t^2} \end{aligned}$$

The rotation matrix corresponds to the following coordinate transformation

$$\begin{aligned} x &= 0.526y_1 + 0.851x_1 \\ y &= 0.851y_1 - 0.526x_1 \end{aligned}$$

from which we can recover a parameterization of the original hyperbola. ◇

Since a rotation affects off-diagonal elements other than the one we are zeroing out, a complete implementation of this approach must apply rotations iteratively, possibly more than once for each off-diagonal element. To understand the nature of the iteration, we recall the concept of matrix norms from Section 4.2.3 of Chapter 4. In particular, the *Frobenius norm* of the $m \times n$ matrix A is defined as

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

The Frobenius norm has the evident property that it is invariant under rotation. Hence, if $C = R^T AR$, then $\|A\|_F = \|C\|_F$. In the remainder of this section, we will use only the Frobenius norm and so will drop the subscript F.

We introduce a measure of how close the matrix A is to being diagonal and consider the quantity

$$\mathrm{off}(A) = \sum_{i \neq j} |a_{ij}|^2$$

It is our plan to reduce this quantity with each rotation. If we can do that, then by repeated rotations we can minimize $\mathrm{off}(A)$, thereby making progress toward diagonalizing the matrix. Note that

$$\|A\|^2 = \mathrm{off}(A) + \sum_i |a_{ii}|^2$$

Let $C = R^T AR$ be a rotation that zeros the ij-element in A. Then the Frobenius norms of A and C are equal. Observing how the diagonal elements a_{ii} and a_{jj} change, we note that

$$\mathrm{off}(C) = \mathrm{off}(A) - 2|a_{ij}|^2 + 2|b_{ij}|^2 = \mathrm{off}(A) - 2|a_{ij}|^2$$

Hence, the quantity off(C) has decreased with twice the square of the entry in A that we zeroed. It follows that repeated rotations, each zeroing an off-diagonal element of largest magnitude, will drive off(A) to zero. For example, parameterizing the parabola $x^2+2xy+y^2+2y-1=0$ in this way will require more than three rotations.

This iterative algorithm for diagonalizing A is numerically stable and has quadratic convergence. For the purpose of backsubstitution, the product of the individual rotation matrices should be accumulated.

Second Method of Conic Parameterization

A *pencil of lines* through a point p is a set of lines each containing p. The geometric idea on which the second parameterization method of conics is based can be stated as follows:

1. Pick a point $p = (u, v)$ on the conic and consider a pencil of lines through p. There is a one-parameter family of lines in the pencil.

2. The line $l(t)$ in the pencil will intersect the conic in p and in one other point $(x(t), y(t))$. This additional point provides the curve parameterization.

Therefore, a point p must be found, and the lines in the pencil must be quantified by a parameter t. For t, there is usually a natural choice. We demonstrate the idea with the circle $x^2 + y^2 - 1$. Thereafter, we show how the geometric idea expresses itself algebraically.

Consider the circle $x^2 + y^2 - 1 = 0$. We choose the point $p = (-1, 0)$ and consider the lines

$$l(t):\ y = tx + t$$

that pass through p; see also Figure 5.4. Then the intersection points of the lines in the pencil and the circle are found by substituting for y and solving for x:

$$x^2(1+t^2) + 2t^2x + t^2 - 1 \;=\; 0$$

hence

$$x \;=\; \frac{-t^2 \pm \sqrt{t^4 + (1-t^2)(1+t^2)}}{1+t^2} \;=\; \frac{-t^2 \pm 1}{1+t^2}$$

Of the two roots, -1 represents the x coordinate of the fixed point p, whereas the other yields the x coordinate of the variable point $x(t)$ in which $l(t)$ also intersects the circle. In this way, we derive the familiar form

$$\begin{aligned} x(t) &= \frac{1-t^2}{1+t^2} \\ y(t) &= \frac{2t}{1+t^2} \end{aligned}$$

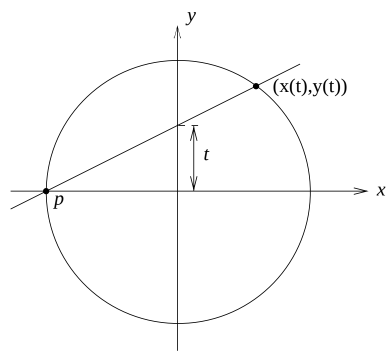

Figure 5.4 Parameterization of the Unit Circle

The procedure has a simple algebraic expression if the point p is chosen with care. To grasp this fact more generally, we pass to homogeneous coordinates (x, y, w).[5] Let

$$a_{11}x^2 + a_{22}y^2 + a_{33}w^2 + 2a_{12}xy + 2a_{13}xw + 2a_{23}yw = 0$$

be the homogenized equation. In general, we expect that none of the quadratic terms will vanish. However, if the conic contains one of the fundamental points, then the equation becomes linear in the corresponding variable. For example, if $(0, 0, 1)$ is a point on the conic, then the equation has no w^2 term. Similarly, if $(0, 1, 0)$ is on the conic, the y^2 term is absent. Thus, when a fundamental point is on the conic, then the conic equation will be linear in one variable, and that variable is then an explicit function of the other two. By considering the other two variables as projective parameter coordinates, therefore, we have a parameterization of the conic.

Example 5.4: Given the circle $x^2 + y^2 - 2xw = 0$ that contains the origin $(0, 0, 1)$, we derive the parameterization:

$$\begin{aligned} x &= s \\ y &= t \\ w &= (s^2 + t^2)/2s \end{aligned}$$

This is a projective parameterization of the projective form of the conic. We set $s = 1$ to obtain affine parameter coordinates, and divide by w to pass

[5] Note that the homogenizing variable w is the third coordinate.

to affine curve coordinates. For the affine form of the circle, therefore, we obtain the affine parameterization

$$\begin{aligned} x(t) &= 2/(1+t^2) \\ y(t) &= 2t/(1+t^2) \end{aligned}$$

◇

Our strategy for parameterizing a conic is to translate or rotate the coordinate system so that the curve will pass through one of the three fundamental points, and then to parameterize the transformed curve. Applying the inverse coordinate transformation, we so obtain a parameterization for the original conic. The structure of the parameterization for conics is thus as follows:

1. If the curve already contains one of the fundamental points $(0, 0, 1)$, $(0, 1, 0)$, or $(1, 0, 0)$, then skip steps 2 and 4.

2. If the curve has a real point at infinity, change the coordinate system such that this point becomes $(0, 1, 0)$. If there is no real point at infinity, find a real point at finite distance and change the coordinate system such that the point becomes $(0, 0, 1)$.

3. Parameterize the curve.

4. Apply the inverse transformation to the parameterization.

Preference is given to finding a real point at infinity because that is a simpler computation.

A point at infinity can be found by setting $w = 0$ in the homogeneous conic equation. The resulting quadratic equation $a_{11}x^2 + 2a_{12}xy + a_{22}y^2 = 0$ is homogeneous in x and y, and has the solution

$$\begin{aligned} x &= -a_{12} \pm \sqrt{a_{12}^2 - a_{11}a_{22}} \\ y &= a_{11} \end{aligned}$$

If the solution is complex, then no real points exist at infinity. This will be the case whenever the discriminant $a_{12}^2 - a_{11}a_{22}$ is less than zero. Otherwise, let $(u, v, 0)$ be the curve point corresponding to the real solution (u, v) of the homogeneous equation. Then the transformation

$$\begin{aligned} x &= x_1 + uy_1 \\ y &= vy_1 \\ w &= w_1 \end{aligned}$$

is nonsingular and brings this point to $(0, 1, 0)$. After dehomogenizing, the transformed conic has the form

$$y_1(cx_1 + d) + q(x_1) = 0$$

where $q(x_1)$ is a quadratic polynomial in x_1. This curve is parameterized as $x_1 = t$, $y_1 = -q(t)/(ct + d)$. Using the transformation equations, a parameterization of the original curve is obtained.

If the curve has no real point at infinity and is not imaginary, then the conic must be an ellipse. A real point on it can be found by locating a point at which one of the partials vanishes, say $f_x = 0$. The point can then be brought to the origin by translation, after which the curve equation is parameterized as described before.

Example 5.5: Consider the conic $x^2+6xy+5y^2-2x-2y-1 = 0$, whose homogeneous form is $x^2+6xy+5y^2-2xw-2yw-w^2 = 0$. The discriminant of $x^2 + 6xy + 5y^2$ is 4, so we expect two real solutions, corresponding to two points at infinity; that is, the conic is a hyperbola. One of the solutions is $(-1, 1)$, corresponding to the point $(-1, 1, 0)$ at infinity. The transformation

$$\begin{aligned} x &= x_1 - y_1 \\ y &= y_1 \end{aligned}$$

maps the (affine) curve to

$$4x_1y_1 + x_1^2 - 2x_1 - 1 = 0$$

which is parameterized as

$$\begin{aligned} x_1(t) &= t \\ y_1(t) &= \frac{1 + 2t - t^2}{4t} \end{aligned}$$

Backtransformation yields, for the original curve, the parameterization

$$\begin{aligned} x(t) &= \frac{5t^2 - 2t - 1}{4t} \\ y(t) &= \frac{1 + 2t - t^2}{4t} \end{aligned}$$

◇

Example 5.6: Consider the conic $x^2 + 4y^2 - 2x - 16y + 13 = 0$. Since $x^2 + 4y^2$ has a negative discriminant, the conic has no real points at infinity; thus, it is either an ellipse or an imaginary ellipse. The partial derivative by x defines the line $2x - 2 = 0$, and the intersection of this line with the conic determines the two points at which the curve has a tangent parallel to the x axis. We substitute 1 for x in the conic to locate these points, obtaining $4y^2 - 16y + 12$; hence, $(1, 3)$ is a point on the conic with tangent parallel to the x axis. We translate the curve by

$$\begin{aligned} x &= x_1 + 1 \\ y &= y_1 + 3 \end{aligned}$$

and obtain

$$x_1^2 + 4y_1^2 + 8y_1 = 0$$

Note that this curve contains the origin, and that its homogeneous form is linear in w. We consider how the lines $x_1 - ty_1 = 0$ intersect this conic. Substituting, we obtain

$$t^2 y_1^2 + 4y_1^2 + 8y_1 = y_1(t^2 y_1 + 4y_1 + 8) = 0$$

Here, $y_1 = 0$ corresponds to the intersection at the origin; hence, the lines intersect the curve at $(0, 0)$ and at

$$\begin{aligned} y_1 &= \frac{-8}{t^2 + 4} \\ x_1 &= \frac{-8t}{t^2 + 4} \end{aligned}$$

This constitutes a parameterization of the translated curve. Translating back, we obtain the parameterization

$$\begin{aligned} x &= \frac{-8t}{t^2 + 4} + 1 = \frac{t^2 - 8t + 4}{t^2 + 4} \\ y &= \frac{-8}{t^2 + 4} + 3 = \frac{3t^2 + 4}{t^2 + 4} \end{aligned}$$

for the original conic. $\diamond$

5.5.2 Quadrics

First Method

As in the case of conics, we can iteratively diagonalize the matrix representation of the quadric using rotations. Let

$$\begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}$$

be the diagonal matrix so obtained. We classify the surfaces first by rank. Here, rank 2 and 1 are not of interest, since in this case the quadric surface consists of two planes.

Rank 4 splits into several cases, according to the signature of the matrix; that is, according to the distribution of signs of the diagonal elements. After suitably renaming variables, and possibly multiplying the conic equation with -1, we have three different cases to distinguish:

1. $\lambda_1, \lambda_2, \lambda_3, \lambda_4 > 0$: The quadric is imaginary. We denote this case by $(+,+,+,+)$.

2. $\lambda_1, \lambda_2, \lambda_3 > 0, \lambda_4 < 0$: The quadric is elliptic. We denote this case by $(+,+,+,-)$.

3. $\lambda_1, \lambda_2 > 0, \lambda_3, \lambda_4 < 0$: The quadric is hyperbolic. We denote this case by $(+,+,-,-)$.

Similarly, for rank 3, we distinguish the cases $(+,+,+,0)$, and $(+,+,-,0)$. The case $(+,+,+,0)$ means that the surface is imaginary. The projective parameterization of the nonimaginary surfaces is given in Table 5.2. In each case, we must multiply the i^{th} coordinates with $\mu_i = 1/\sqrt{|\lambda_i|}$. Note that the rank 3 surfaces are cones and cylinders.

Second Method

The second method for parameterizing conics also generalizes to quadrics: A real point is picked on the surface, and a pencil of lines through this point is considered. Again, the lines intersect the quadric in one additional point, and we obtain in this way a surface parameterization. Since the lines are in three-dimensional space, each member of the pencil must be fixed by two independent parameters.

The algebraic method is closely analogous to the one for conics. We wish to move a real point of the surface to one of the fundamental points of projective three-dimensional space — that is, to $(0,0,0,1)$, $(0,0,1,0)$, $(0,1,0,0)$, or

Signature	Parametric Form	
$(+,+,+,-)$	$x = r^2 - s^2 - t^2$	$y = 2rs$
	$z = 2rt$	$w = r^2 + s^2 + t^2$
$(+,+,-,-)$	$x = r^2 - s^2 + t^2$	$y = 2rs$
	$z = 2rt$	$w = r^2 + s^2 - t^2$
$(+,+,-,0)$	$x = r^2 - s^2 + t^2 - 2rt$	$y = 2rt - r^2 - s^2 - t^2$
	$z = 2sr - 2st$	$w = 1$

Table 5.2 Projective Parameterization of Quadric Surfaces

$(1, 0, 0, 0)$. Correspondingly, the equation simplifies with one of the quadratic terms vanishing in the homogeneous form. Thereafter, the parameterization proceeds as in the case of conics. For example, with the surface passing through $(0, 1, 0, 0)$, its equation is linear in y. Hence, choosing $x = s$ and $z = t$ expresses y as a rational function of s and t.

A real surface point at infinity is found by investigating the homogeneous equation at $w = 0$. The substitution $w = 0$ gives a homogeneous form that describes a conic. This is the conic in which the quadric intersects the plane at infinity. This conic may have real points, found as described previously, and any such point can then be moved to $(0, 1, 0, 0)$ by a coordinate transformation. Alternatively, we may deal with a closed surface (i.e., the ellipsoid), in which case we find a real point by locating where on the surface two of its partial derivatives vanish simultaneously. The details are quite straightforward.

5.5.3 Cubic Curves

Not all irreducible cubic curves have a rational parametric form. Those that do have a singular point that must be a double point. From a geometric point of view, the rational parameterization is analogous to conic parameterization: We select the double point on the cubic and consider a pencil of lines through it. Each line in the pencil must intersect the cubic in only one additional point. By parameterizing the pencil, we thus can parameterize the cubic. This idea is illustrated in Figure 5.5 for the curve $y^2 - x^2 - x^3 = 0$, whose double point is the origin. Thus, this curve is parameterized by the pencil of lines $y = tx$.

We describe an algorithm for parameterizing a cubic $f(x, y) = 0$. Using several birational transformations, the algorithm brings the cubic $f(x, y) = 0$ into the form

$$y_2^2 = g(x_2)$$

where the polynomial $g(x_2)$ has degree 4. If $g(x_2)$ has a double root, then

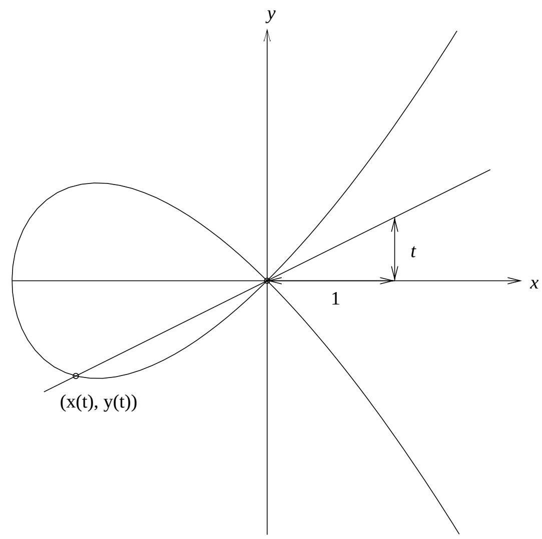

Figure 5.5 Parameterizing a Singular Cubic Curve

the cubic f is parameterizable; otherwise, a rational parametric form does not exist. As we will see, the parameterization process makes use of the parameterization algorithm for conics.

The curve transformations are as follows. First, we find a real point of the curve at infinity and bring it to $(0, 1, 0)$. This new coordinate system will be named (x_1, y_1). A real point at infinity must always exist, and when this point is $(0, 1, 0)$, then the cubic will not have a y_1^3 term. To understand this observation, we substitute $(0, 1, 0)$ into the homogeneous form F of the cubic. We obtain a value equal to the coefficient of the y_1^3 term in F. So, if $(0, 1, 0)$ is a curve point, the coefficient of y_1^3 must vanish.

Second, possibly after multiplying with a linear polynomial in x_1, the curve equation is changed to $y_2^2 - g(x_2) = 0$, where y_2 has the form $h(x_1)y_1 + k(x_1)$, where $h(x_1)$ is at most linear in x_1, and $k(x_1)$ is at most quadratic. Finally, if $g(x_2)$ has the double root α, then we let $y_3 = y_2/(x_2 - \alpha)$ and parameterize the resulting conic in y_3 and x_3. The conic parameterization, in turn, yields a parameterization for the cubic.

Consider a cubic $f(x, y)$ assuming that it is *regular* in both x and y; that is, both the x^3 and the y^3 term are present. To find a real point at infinity, we find a real root of the polynomial formed by all terms of degree 3 in f. Since this polynomial is a cubic homogeneous form, there is always a real root, say $(-v, u)$. We thus can write f as

$$f = (ux + vy)f_2(x, y) + g_2(x, y)$$

where f_2 is homogeneous of degree 2, and g_2 is at most of degree 2. The needed root, $(-v, u)$, could be found after substituting 1 for y, using a numerical subroutine or the Cardano formulae. We transform f by the nonsingular transformation

$$\begin{aligned} x &= x_1 - v y_1 \\ y &= u y_1 \end{aligned}$$

Then the transformed cubic can be written as

$$h_1(x_1)y_1^2 + h_2(x_1)y_1 + h_3(x_1) = 0$$

where h_i is a polynomial of degree at most i. By multiplication with $4h_1(x_1)$, this equation can be rewritten as

$$(2h_1(x_1)y_1 + h_2(x_1))^2 = (h_2(x_1))^2 - 4h_1(x_1)h_3(x_1)$$

Setting $y_2 = 2h_1(x_1)y_1 + h_2(x_1)$ and $x_2 = x_1$, we thus obtain

$$y_2^2 = h_4(x_2)$$

where h_4 has degree 4 or less.

We investigate the roots of h_4. If there is at least one double root α, then we can set $y_3 = y_2/(x - \alpha)$ and so obtain the conic

$$y_3^2 = q(x_2)$$

This conic is parameterized, and, by backsubstitution, a parameterization of the original cubic is obtained. Note that if α is complex, then its conjugate $\tilde{\alpha}$ is also a double root and we set $y_3 = y_2/((x - \alpha)(x - \tilde{\alpha}))$. Thus a parameterization with real coefficients is possible. However, if h_4 has no multiple roots and has degree 3 or 4, then the curve $y_3^2 = h_4(x_1)$ has genus 1 and does not possess a rational parameterization. Since the original cubic f has been mapped to this curve birationally, it follows that f cannot be parameterized and is a nonsingular cubic.

Example Parameterization of a Cubic

Consider the cubic

$$f = 28y^3 + 26xy^2 + 7x^2y + x^3/2 + 28y^2 + 16xy + 7y + 3x/2$$

The degree form is $28y^3+26xy^2+7x^2y+x^3/2$ and has the root $(-2,1)$. That is, we can write

$$f = (x+2y)(14y^2+6xy+x^2/2)+28y^2+16xy+7y+3x/2$$

We substitute

$$\begin{aligned} x &= x_1 - 2y_1 \\ y &= y_1 \end{aligned}$$

to obtain

$$4(x_1-1)y_1^2+4(x_1^2+4x_1+1)y_1+(x_1^3+3x_1)/2=0$$

After multiplication with (x_1-1), we obtain the equivalent form

$$4(x_1-1)^2y_1^2+4(x_1-1)(x_1^2+4x_1+1)y_1+(x_1^4-x_1^3+3x_1^2-3x_1)/2=0$$

We set $y_2 = 2(x_1-1)y_1+x_1^2+4x_1+1$ and obtain

$$y_2^2 = (x_1^4+17x_1^3+33x_1^2+19x_1+2)/2$$

The right-hand side has the double root $x_1=-1$, so we set $y_3 = y_2/(x_1+1)$ to obtain

$$2y_3^2 = x_1^2+15x_1+2$$

This is a conic with the parameterization

$$\begin{aligned} x_1 &= \frac{t^2-2}{2t+15} \\ y_3 &= -\frac{t^2+15t+2}{\sqrt{2}(2t+15)} \end{aligned}$$

Recalling the substitution $y_3 = y_2/(x_1+1)$, we now obtain

$$y_2 = -\frac{(t^2+15t+2)(t^2+2t+13)}{\sqrt{2}(2t+15)^2}$$

Since $y_2 = 2(x_1 - 1)y_1 + x_1^2 + 4x_1 + 1$, we obtain next

$$\begin{aligned} y_1 &= -[(\sqrt{2}+1)t^4 + (8\sqrt{2}+17)t^3 + (60\sqrt{2}+45)t^2 + (44\sqrt{2}+199)t \\ &\quad +(109\sqrt{2}+26)] \,/\, [\sqrt{2}(4t^3 + 22t^2 - 128t - 510)] \end{aligned}$$

Numerator and denominator have the common root $1 - 3\sqrt{2}$. Thus, the parametric expression for y_1 simplifies to

$$y_1 = \frac{(\sqrt{2}+1)t^3 + (6\sqrt{2}+12)t^2 + (30\sqrt{2}+21)t + (11\sqrt{2}+40)}{\sqrt{2}(4t^2 - (12\sqrt{2}-26)t - (90\sqrt{2}+30))}$$

From this parameterization, we finally obtain the parameterization of the original curve.

As an example of a nonsingular cubic, consider $y^2 - x^3 + x = 0$. It is already in the form $y_2^2 = x_2^3 - x_2 = h_4(x_2)$. Here, h_4 is of degree 3 with the distinct roots -1, 0, and 1, so the curve does not have a rational parametric form.

5.5.4 Monoids

A curve of degree n with a point of multiplicity $n-1$ is called a *monoid*. Conics trivially are monoids, as are cubic curves possessing a double point (i.e., singular cubics). By Bezout's theorem, a line through an $(n-1)$-fold point p intersects the curve in at most one additional point. Hence, a pencil of lines through p can be used to parameterize the monoid. The parameterization of conics and cubics has followed this strategy.

When the curve point of multiplicity $(n-1)$ is the origin, then the equation of the monoid has the form

$$f(x,y) = h_n(x,y) - h_{n-1}(x,y) = 0$$

where h_n is homogeneous of degree n and h_{n-1} is homogeneous of degree $n-1$. The curve parameterization is then simply

$$x(s,t) = s\,\frac{h_{n-1}(s,t)}{h_n(s,t)} \qquad\qquad y(s,t) = t\,\frac{h_{n-1}(s,t)}{h_n(s,t)}$$

This parameterization is projective, since it is derived from the pencil of lines

$$\begin{aligned} x(\lambda) &= s\lambda \\ y(\lambda) &= t\lambda \end{aligned}$$

We can make it an affine parameterization by choosing $s = 1$ or $t = 1$. For $s = 1$, we have

$$y = tx,$$

and, for $t = 1$,

$$x = sy.$$

Example 5.7: Consider the monoid

$$x^4 - 3x^3y + x^2y^2 + 2y^4 - x^3 - 3x^2y + y^2x = 0$$

Here, $h_4 = x^4 - 3x^3y + x^2y^2 + 2y^4$ and $h_3 = x^3 + 3x^2y - y^2x$. Choosing $s = 1$, we have $h_4(1,t) = 1 - 3t + t^2 + 2t^4$ and $h_3(1,t) = 1 + 3t - t^2$; hence, we obtain the parametric form

$$\begin{aligned} x(t) &= \frac{1 + 3t - t^2}{1 - 3t + t^2 + 2t^4} \\ y(t) &= \frac{t(1 + 3t - t^2)}{1 - 3t + t^2 + 2t^4} \end{aligned}$$

◇

A surface of degree n is a *monoid* or *monoidal surface* if it contains a point of multiplicity $n - 1$. Monoidal surfaces include all quadrics, any cubic surface with a double point, any quartic surface with a triple point, and so on. When the $(n - 1)$-fold point is at the origin, the equation of such a surface becomes

$$f(x, y, z) = h_n(x, y, z) - h_{n-1}(x, y, z) = 0$$

where h_n has degree n and h_{n-1} has degree $n-1$. The surface is parameterized by a pencil of lines through the origin. Each line in the pencil is determined by a point (r, s, t) of the projective plane and is given by the (parametric) equations

$$\begin{aligned} x(\lambda) &= r\lambda \\ y(\lambda) &= s\lambda \\ z(\lambda) &= t\lambda \end{aligned}$$

Therefore, the projective form of the surface parameterization is simply

$$x(r, s, t) = r\,\frac{h_{n-1}(r, s, t)}{h_n(r, s, t)}$$

$$\begin{aligned} y(r,s,t) &= s\,\frac{h_{n-1}(r,s,t)}{h_n(r,s,t)} \\ z(r,s,t) &= t\,\frac{h_{n-1}(r,s,t)}{h_n(r,s,t)} \end{aligned}$$

and can be changed to the familiar affine parameterization by setting, for instance, $r = 1$.

Example 5.8: Consider the unit sphere

$$x^2 + y^2 + (z-1)^2 - 1 = 0$$

which contains the origin. We have $h_2(x,y,z) = x^2+y^2+z^2$ and $h_1(x,y,z) = 2z$. Hence, this sphere is (projectively) parameterized by

$$\begin{aligned} x(r,s,t) &= \frac{2rt}{r^2+s^2+t^2} \\ y(r,s,t) &= \frac{2st}{r^2+s^2+t^2} \\ z(r,s,t) &= \frac{2t^2}{r^2+s^2+t^2} \end{aligned}$$

as is readily verified. An affine parameterization can be derived by setting one of the three parameters to 1. $\diamond$

5.5.5 Parametric Domains

We mentioned previously that there is a rich literature on parametric curves and surfaces. To be more precise, the literature on that subject concentrates on *patches* of parametric curves and surfaces; that is, only a finite part of the curve or surface is considered. Typically, the patch is defined by restricting the parameter(s) to a *domain.* In the case of curves, the domain might be the interval $[0,1]$; in the case of surfaces, the domain might be the unit square $[0,1] \times [0,1]$.

So far, we have discussed parameterizing curves and surfaces without regard to how the parameterization might be used. For example, if we consider a patch on the surface just parameterized, we may want to adjust the parameterization such that the patch is defined over a standard domain. We therefore consider the following problem.

> **Problem**
> Given a parameterized surface, and given four distinct surface points, by their parametric coordinates; reparameterize the surface, such that the parametric coordinates of the four given points are the corners of the unit square.

The reparameterization can be done based on a projective parameterization.

From the given projective parameterization (r, s, t), we derive another projective parameterization (u, v, w) such that the four corners are mapped as desired. Let the four corners be (r_i, s_i, t_i), in clockwise order. We seek a linear transformation A relating the two parameterizations as follows:

$$\rho \begin{pmatrix} r \\ s \\ t \end{pmatrix} = A \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$

such that (r_1, s_1, t_1) is mapped to $(0, 0, 1)$, (r_2, s_2, t_2) is mapped to $(1, 0, 1)$, and so on. Note that, since (r, s, t) and $(\rho r, \rho s, \rho t)$ determine the same point on the projective plane, the proportionality factor ρ is needed. From projective geometry, we know that such a linear map exists, provided that no three of the points (r_i, s_i, t_i) are collinear. Thereafter, the (u, v, w) parameterization is dehomogenized by setting $w = 1$.

We formulate a system of linear equations determining A. The unknowns are the coefficients a_{jk} of A, and the proportionality factors ρ_i. For $i = 1, 2, 3, 4$, we write

$$\rho_i \begin{pmatrix} r_i \\ s_i \\ t_i \end{pmatrix} = A \begin{pmatrix} u_i \\ v_i \\ w_i \end{pmatrix} \tag{5.2}$$

We obtain 12 linear equations in 13 unknowns. Solving the system, we obtain A, and hence the new parameterization with the required domain.

Example 5.9: Consider the unit sphere parameterized as before:

$$\begin{aligned} x(r, s, t) &= \frac{2rt}{r^2 + s^2 + t^2} \\ y(r, s, t) &= \frac{2st}{r^2 + s^2 + t^2} \\ z(r, s, t) &= \frac{2t^2}{r^2 + s^2 + t^2} \end{aligned}$$

We wish to map a surface patch with the parametric corner coordinates $(0, 2, 1)$, $(1, 4, 1)$, $(3, 5, 1)$, and $(4, 1, 1)$; see also Figure 5.6. To find the appropriate projective transformation, we formulate the linear equations (5.2)

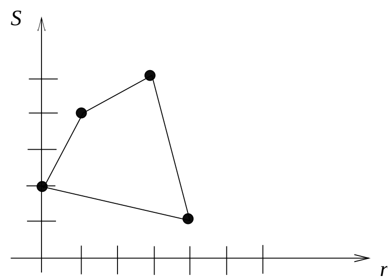

Figure 5.6 Patch Corners in Parameter Space

and obtain

$$\begin{aligned} 0 &= a_{13} \\ 2\rho_1 &= a_{23} \\ \rho_1 &= a_{33} \\ &\vdots \\ \rho_4 &= a_{31} + a_{33} \end{aligned}$$

Note that the a_{jk} are the entries of A. The solution yields the matrix

$$A = \lambda \begin{pmatrix} 4 & 5 & 0 \\ -5 & 14 & 6 \\ -2 & 2 & 3 \end{pmatrix}$$

Note that λ is a free constant. So, we can reparameterize the sphere by substituting $4u + 5v$ for r, $-5u + 14v + 6w$ for s, and $-2u + 2v + 3w$ for t. After dehomogenizing with $w = 1$, the resulting parameterization maps the corners of the domain $[0, 1] \times [0, 1]$ as desired. $\diamond$

5.6 Conversion from Parametric to Implicit Form

Classical elimination theory provides tools for converting from rational parametric representations to implicit representations. Briefly, if a curve is given as

$$\begin{aligned} x(t) &= p(t)/r(t) \\ y(t) &= q(t)/r(t) \end{aligned}$$

then the pair of polynomial equations, obtained by clearing the denominator $r(t)$, describes the curve

$$\begin{aligned} x \cdot r(t) - p(t) &= 0 \\ y \cdot r(t) - q(t) &= 0 \end{aligned}$$

These are polynomial equations in the variables x, y, and t. If t is eliminated from them, a single equation in x and y is obtained that is the implicit curve equation. Similarly, one eliminates the two parameters s and t from three polynomial equations obtained from a parametric surface representation, thus implicitizing a surface.

5.6.1 Resultants

One method for eliminating a variable from two polynomial equations is by forming the *resultant*. In the simplest case, there is only one variable to be eliminated; and thus we are, in effect, testing whether the two polynomials have a common root. We discuss this case in some detail. Let

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_0$$

and

$$g(x) = b_m x^m + b_{m-1} x^{m-1} + \ldots + b_0$$

be two univariate polynomials, of degree n and m. Clearly, f and g have a common root iff there are polynomials h_f and h_g of degree less than m and n, respectively, such that

$$f(x) h_f(x) = g(x) h_g(x) \tag{5.3}$$

We set

$$h_f(x) = u_{m-1} x^{m-1} + \ldots + u_1 x + u_0$$

and

$$h_g(x) = v_{n-1} x^{n-1} + \ldots + v_1 x + v_0$$

Then the coefficients u_k and v_k can be determined by symbolically multiplying out equation (5.3). The result is a polynomial in x all of whose coefficients must vanish. Each coefficient, in turn, is a linear form in the unknowns u_k and v_k. By setting the x coefficients to zero, we obtain a system of linear equations in u_k and v_k, and this system has a nontrivial solution iff equation (5.3) has a nontrivial solution. Now equation (5.3) has a nontrivial solution iff the linear system is underconstrained; that is, iff the system's determinant is zero. The system's determinant has the following form:

$$\begin{vmatrix} a_n & a_{n-1} & \cdots & a_0 & 0 & \cdots & 0 \\ 0 & a_n & \cdots & a_1 & a_0 & \cdots & 0 \\ \vdots & & \ddots & & & \ddots & \\ 0 & \cdots & 0 & a_n & a_{n-1} & \cdots & a_0 \\ b_m & b_{m-1} & \cdots & b_0 & 0 & \cdots & 0 \\ 0 & b_m & \cdots & b_1 & b_0 & \cdots & 0 \\ \vdots & & \ddots & & & \ddots & \\ 0 & \cdots & 0 & b_m & b_{m-1} & \cdots & b_0 \end{vmatrix}$$

So, f and g have a common solution iff this determinant is zero. The determinant is called the *Sylvester resultant*, and we will denote it $Res_x(f, g)$.

In principle, the resultant can be applied to multivariate polynomials. A main variable x is identified, and the coefficients a_k and b_k are now polynomials in the remaining variables. In this case, a zero resultant does not necessarily imply a common solution to the two polynomials, since it is possible that the two lead coefficients a_n and b_m have a common solution. In that case, the resulting polynomial $Res_x(f, g)$ has additional factors identifying the common roots of the lead coefficients. Summarizing, the following theorem can be proved.

Theorem
Let $f(x_0)$ and $g(x_0)$ be multivariate polynomials of degree n and m, respectively, with coefficients that are polynomials in the variables $x_1, ..., x_n$. Then $Res_x(f, g) = 0$ iff there is a common solution of f and g, or the leading coefficients of f and g vanish simultaneously, or the coefficient polynomials of f or of g have a common root.

Consider the bivariate polynomials $f(x, y) = xy^2 - x^2y + y^2 + x^2 + 1$ and $g(x, y) = x^2y^2 - y^2 - 3xy + x$. We consider them polynomials in y with coefficients that are polynomials in x, and obtain as resultant

$$Res_y(f, g) \quad = \quad \begin{vmatrix} x+1 & -x^2 & x^2+1 & 0 \\ 0 & x+1 & -x^2 & x^2+1 \\ x^2-1 & -3x & x & 0 \\ 0 & x^2-1 & -3x & x \end{vmatrix} \qquad \text{so that}$$

$$Res_y(f, g) = (x + 1)(x^7 - 3x^6 + x^5 + 2x^4 + 3x^3 + 11x^2 + x + 1)$$

When formulating f and g as polynomials in y, the lead coefficients are $x+1$ and x^2-1, respectively. Both vanish simultaneously for $x=-1$, as reflected in the presence of the factor $x+1$.

Geometrically, the Sylvester resultant constitutes an orthographic projection along the axis of the variable that is eliminated. The following consideration shows that some extraneous factors have a geometric significance. Consider three polynomial equations in three variables — say, $f(x,y,z)=0$, $g(x,y,z)=0$, and $h(x,y,z)=0$. The three equations define three surfaces, and the common solutions to f, g, and h are the points at which all three surfaces intersect. We eliminate first z, obtaining two equations in x and y — say, $f_1(x,y) = Res_z(f,g)$ and $g_1(x,y) = Res_z(g,h)$. Note that the curve $f_1 = 0$ contains the projection of the space curve that is the intersection $f=0 \cap g=0$. Similarly, $g_1 = 0$ contains the projection of the intersection $g=0 \cap h=0$. A common intersection of the three surfaces must also be an intersection of the two plane curves f_1 and g_1. However, if there are two points $p=(a,b,c)$ and $q=(a,b,d)$ in 3-space, where $c \neq d$, and p is on the intersection $f \cap g$ while q is on the intersection $g \cap h$, then $p' = (a,b)$ is an intersection of $g_1 = 0$ with $h_1 = 0$ but is not a common intersection of the three surfaces. These "phantom" intersections must give rise to extraneous factors when we eliminate one of the variables from f_1 and g_1.

5.6.2 Implicitization of Curves and Surfaces

The resultant provides us with an algorithm to convert parametric curves and surfaces into implicit form. In the case of curves, we need to compute

$$f(x,y) = Res_t(xr(t)-p(t),\ yr(t)-q(t))$$

For example, recalling the parameterization of the unit circle (Table 5.1), we have

$$Res_t(x(1+t^2)-(1-t^2),\ y(1+t^2)-2t) \ = \ 4(x^2+y^2-1)$$

Here, no extraneous polynomial factors appear.

In the case of surfaces, we must eliminate two variables in succession. For example, the sphere that we parameterized as a monoid, with $r=1$, requires dealing with the equations

$$x(1+s^2+t^2)-2t = 0 \qquad (5.4)$$
$$y(1+s^2+t^2)-2st = 0 \qquad (5.5)$$
$$z(1+s^2+t^2)-2t^2 = 0 \qquad (5.6)$$

We begin by eliminating s from equations (5.4) and (5.5), and also from equations (5.5) and (5.6). This yields the equations

$$\begin{aligned} 4t^2(y^2 + t^2x^2 + x^2 - 2tx) &= 0 \\ 4t^2(t^2z^2 + z^2 - 2t^2z + t^2y^2) &= 0 \end{aligned}$$

After dropping the common factor $4t^2$, we now eliminate t and obtain

$$(x^2 + y^2 + z^2 - 2z)(y^6 + y^4(x^2 + z^2) - 2y^2z(y^2 + 2x^2) + 4x^2z^2) = 0$$

Note that the second factor is extraneous.

We observe that surface implicitization using the Sylvester resultant is not an attractive method, since even in such simple examples complicated extraneous factors are generated. Since polynomial factorization is a difficult problem, it is not always easy to recognize the extraneous factors and to eliminate them. Moreover, the Sylvester resultant requires forming large matrices for higher-degree surfaces that can be costly to evaluate. We will discuss some alternatives in Sections 7.5.1 and 7.8.3 in Chapter 7.

5.7 Edge Identification

We have divided the representation of edges and faces into two parts: A description of the carrier (a space curve or surface) and a description of the boundary delimiting the area or interval of interest on the carrier. The preceding material in this chapter has dealt with techniques for representing the carrier, and has described elementary methods for manipulating these representations. We now turn to the boundary specification for edges. If curved edges are not specified carefully, the boundary description of objects could contain ambiguities.

In favorable cases, the carrier of the edge is a space curve that possesses a parameterization. In that case, one may represent the carrier parametrically and identify the edge by giving an interval of parameter values. This identifies the edge unambiguously. In general, however, the carrier is not parameterizable and so must be defined as the intersection of surfaces. In that case, the identification of the edge on the carrier is more delicate.

As a segment of a space curve, an edge boundary consists simply of the two bounding vertices; that is, of the two curve points. However, the global geometry of the carrier may be such that the two points do not identify a curve segment uniquely, so that additional information will be needed. In the following discussion, we assume that the edge carrier has been specified by an intersecting pair of surfaces.

5.7.1 Topological Aspects

Recall from Chapter 2, Section 2.4, the definition of a topologically valid solid. The definition considered two separate aspects:

1. The topological structure of solids was characterized abstractly, in the simplest case as a connected 3-manifold with compact boundary.

2. The relationship between the solid and the surrounding space was characterized by considering how the solid is embedded in three-dimensional space.

If we begin with a geometric description of the carrier, the specification of the edge as a segment on that carrier seems to be simply a matter of identifying a startpoint and an endpoint. However, there are complications.

If the curve is closed, then two points on it partition the curve into two segments, and we need to know which of the two segments is the edge. This problem did not occur with straight lines, for at most one of the line segments is finite.[6] This motivates orienting the carrier. If the carrier is a simple closed curve, distinct start and end vertices will specify a segment on it unambiguously.

If the space curve contains singularities, then it does not need to be homeomorphic to a circle. In that case, two points may partition the curve into more than two segments, so there could be several segments, each oriented from the start to the end vertex. An example is shown later in Figure 5.12. Here, we ask whether some segments can be ruled out because they would lead to geometric or topological inconsistencies later. We will show that a global resolution of edge ambiguities cannot be guaranteed.

The constructions demonstrate the need to identify segments of space curves by more information than just bounding vertices and orientation. Two methods have been proposed. One method uses an auxiliary vertex placed at the interior of the edge; the other method provides additional directional information at the two bounding vertices.

5.7.2 Edge Orientation

In Chapters 2 and 3, we oriented edges so that there could be a reference direction, giving meaning to concepts such as left and right adjacent faces. In the linear case, edges have distinct vertices u and v, and the edge (u, v) may simply be considered oriented from u to v. When extending this approach to curved edges, we must make some modifications: Since a curved edge may be closed, the specification (u, v) cannot imply an orientation. A subdivision of such edges is necessary to define an edge orientation at the same time, using this technique. A minimum of three vertices is needed on a closed curve. A

[6]In projective space, a line is a closed curve. Two distinct points define two segments, but at most one is finite. This segment must be chosen because we assume that the boundary of solids is compact.

disadvantage of this approach is that we must traverse an edge, say (u, v), before we can decide whether the initial direction of traversal was the correct one. This is unsatisfactory in general.

A better method for orienting edges is to orient the carrier directly. Let f and g be two surfaces whose intersection contains the edge. Let ∇f and ∇g be the surface gradients. We consider orienting the curve locally at the point p by the directed tangent

$$\mathbf{t}(p) = \nabla f(p) \times \nabla g(p)$$

As long as p is not singular on f or on g, and f and g are not tangent to each other at p, the vector $\mathbf{t}(p)$ exists. Let us call this convention the *cross-product method*.

The cross-product method has several properties that complicate its use. In particular,

$$\nabla f \times \nabla g = -(\nabla g \times \nabla f) = \nabla(-g) \times \nabla f$$

So, we must distinguish between f and $-f$, and hence must adopt the surface-orientation conventions of Section 3.2 in Chapter 3. Now we must give the two surfaces in order. We propose to fix this order *implicitly* as

(left face, right face)

declaring the carrier orientation to be $\nabla f \times \nabla g$, where $f = 0$ is the carrier of the left face, and $g = 0$ is the carrier of the right face, and both f and g have been oriented correctly.

However, this convention gives the expected results only when the angle between the face normals at p is acute. Otherwise, the order should be (right face, left face); see also Figure 5.7. Therefore, we specify the orientation by explicitly annotating the edge with the surface pair, ordered depending on the angle of intersection.

The angle between two curved faces varies along the edge, and our convention is defective if it changes from acute to obtuse, or vice versa, because then the carrier orientation as described by the gradient pair should be reversed. An example is shown in Figure 5.8, with the intersection of $f: \ z = 0$ and $g\colon \ z + y^2 - x^2 - x^3 = 0$, a plane curve, oriented uniformly as $\nabla f \times \nabla g$. The orientation reverses at the singularity at the origin where ∇f and ∇g are collinear. It is not difficult to see that a surface intersection curve must have a singularity at every point p at which the surface gradients are collinear. Hence, it is useful to require that there be no singularities in an edge interior. Thus, we require a vertex at every singular curve point that is part of the surface of the object that we wish to represent.

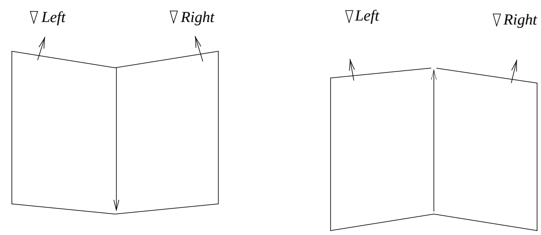

Figure 5.7 Implicit Edge Orientation as Left Face × Right Face

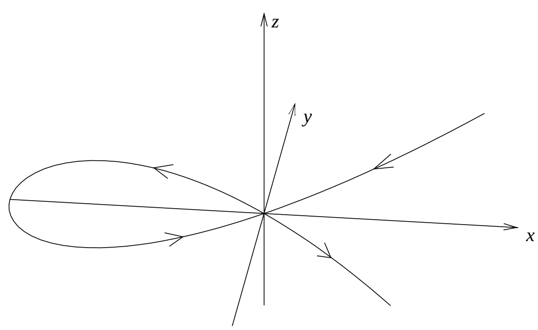

Figure 5.8 Intersection of $f\colon\ z = 0$ and $g\colon\ z + y^2 - x^2 - x^3 = 0$, Oriented as $\nabla f \times \nabla g$

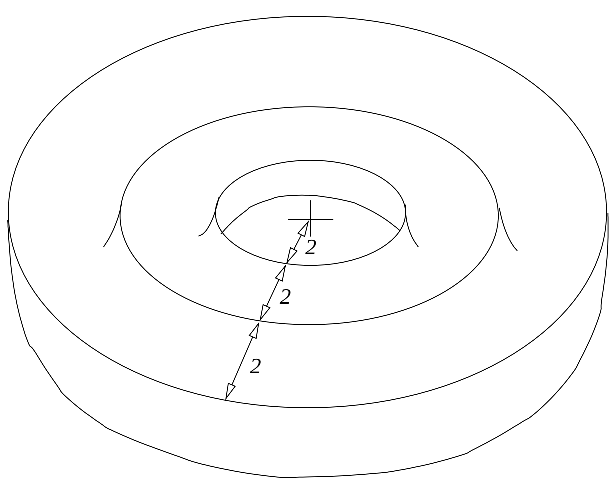

Figure 5.9 Grooved Toroidal Object

If the carrier is a parameterized space curve, then its parameterization implies an orientation. Most of the difficulties in specifying the edge unambiguously disappear in this case, for we consider the edge to be the curve segment defined by an interval of parameter values. Since not all algebraic space curves can be parameterized, however, this method is of limited utility.

5.7.3 Singularities on Edges

If an edge (u, v) is assumed to be oriented by the cross-product method, subject to the details just described, it is still possible that there are two or more curve segments that are both oriented from u to v. In this case, the edge may have been ambiguously defined, since we do not know locally which of the segments (u, v) is intended.

Conceivably, some of the segments cannot be used because of other properties of the boundary. For example, by choosing a particular segment e, we might not be able to find a consistent boundary for some of the faces. Possibly, then, such a local ambiguity could be resolved by global geometric properties of the data structure. We demonstrate now that global properties need not resolve ambiguities when we use implicit algebraic surfaces.

We assume that there are no curve singularities in the interior of edges, that every edge cycle contains at least three vertices, and that the segment orientation so implied is consistent with the local curve orientation by the cross-product method. With all these conventions, we now construct an ambiguous boundary representation.

We unambiguously construct an object with CSG operations. Then, we give a boundary representation for it and show that there is a second interpretation that defines a different solid. In the CSG definition, each primitive volume is specified by an implicit algebraic equation $f = 0$, and is the closure of the set of all points p for which $f(p) < 0$. Thus, we deal with regular sets, albeit not always of finite volume. Specifically, we use the *Cartesian* cylinder $C: \; y^3 + z^3 - 6yz = 0$, and the standard CSG primitives.

We construct the grooved toroidal object shown in Figure 5.9 by subtracting tori T_2 and T_3 from the halved torus $T_1 \cap H$, where H is the half-space $y \leq 0$. The corresponding CSG expression is $((T_1 \cap H) - T_2) - T_3$. The tori dimensions are as indicated in the figure. Next, we take the cylinder C, described previously and shown in Figure 5.10. The surface orientation of C is as indicated by the gradient vectors drawn in the figure. The intersection $(((T_1 \cap H) - T_2) - T_3) \cap C$ is as shown in Figure 5.11, and is the final object. We give a boundary description of the object in Tables 5.3–5.5. The boundary description could be the result of a conversion algorithm translating CSG trees to boundary representations, or the description could have been constructed directly from Figure 5.11 by an unwary designer.

In the description, we assume that the edges are oriented as specified by the vertex pair written, and this is consistent with the carrier orientation by the cross-product of left-face gradient with right-face gradient, and by the

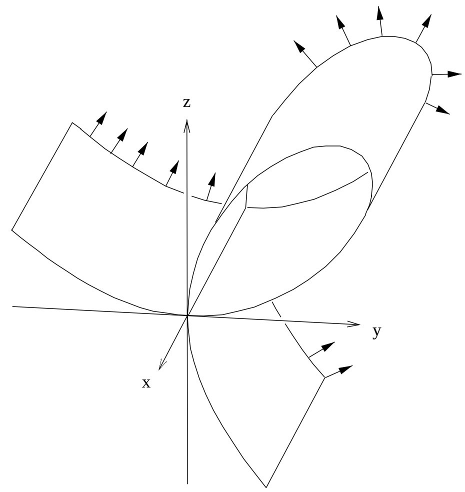

Figure 5.10 The Cartesian Cylinder C

cross-product of the surface gradients as indicated. Edge e_1, then, is oriented from vertex v_1 to vertex v_3. The opposite direction is indicated by negating the edge symbol; for instance, $-e_1$ denotes the edge e_1 in opposite direction.

We consider which curve segment constitutes an edge in the description given in the tables. The complete intersection curve of the torus T_1 with C is shown in Figure 5.12. Based on local information, the edge (v_6, v_4) can lie in one of four directions at v_6. One, in the $-x$ direction, must be excluded, since it does not directly connect to v_4. Another one, in the x direction, cannot

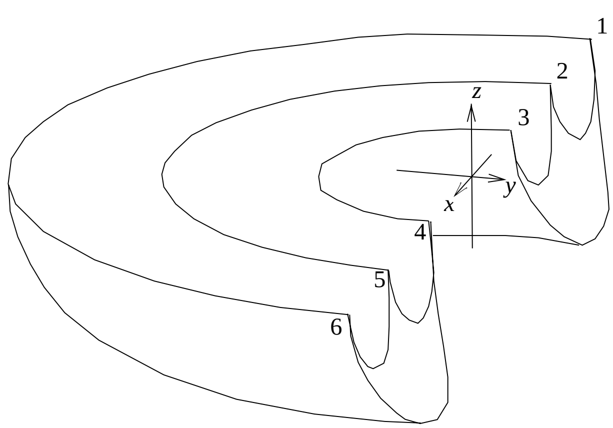

Figure 5.11 Object $(((T_1 \cap H) - T_2) - T_3) \cap C$

Vertex Coordinates	
v_1:	$(-6, 0, 0)$
v_2:	$(-4, 0, 0)$
v_3:	$(-2, 0, 0)$
v_4:	$(2, 0, 0)$
v_5:	$(4, 0, 0)$
v_6:	$(6, 0, 0)$

Vertex Incidences	
v_1:	$(e_1, -e_3, e_7)$
v_2:	$(e_3, -e_2, e_8)$
v_3:	$(e_2, -e_1, e_9)$
v_4:	$(e_5, -e_9, -e_4)$
v_5:	$(e_6, -e_8, -e_5)$
v_6:	$(e_4, -e_7, -e_6)$

Table 5.3 Vertex Tables of $(((T_1 \cap H) - T_2) - T_3) \cap C$

be correct, since then the left and right faces of the edge would be incorrectly situated. They also would require a different edge orientation by the cross-product convention. The remaining two directions give consistent interpretations with the convention of outward-pointing normals, the topology of the boundary description, and the curve orientation by the cross-product method. For global topological consistency, the direction choices must agree at all vertices. We verify that there are two consistent interpretations of the boundary description. The second interpretation is shown in Figure 5.13. Note that the two interpretations are not congruent to each other and have different volumes.

Face Equation	
a:	$(x^2 + y^2 + z^2 - 4)^2 + 32(z^2 - x^2 - y^2 - 4) + 256 = 0$
b:	$-(x^2 + y^2 + z^2 - 1)^2 - 18(z^2 - x^2 - y^2 - 1) - 81 = 0$
c:	$-(x^2 + y^2 + z^2 - 1)^2 - 50(z^2 - x^2 - y^2 - 1) - 625 = 0$
d:	$y^3 + z^3 - 6yz = 0$
e:	$y^3 + z^3 - 6yz = 0$

Face Boundary	
a:	$(-e_1, e_7, e_4, -e_9)$
b:	$(-e_2, e_9, e_5, -e_8)$
c:	$(-e_3, e_8, e_6, -e_7)$
d:	(e_1, e_2, e_3)
e:	$(-e_4, -e_6, -e_5)$

Table 5.4 Face Tables of $(((T_1 \cap H) - T_2) - T_3) \cap C$

Edge	Incident Vertices	Left, Right Face	Carrier Orientation
e_1	(v_1, v_3)	(a, d)	$a \times d$
e_2	(v_3, v_2)	(b, d)	$b \times d$
e_3	(v_2, v_1)	(c, d)	$c \times d$
e_4	(v_6, v_4)	(e, a)	$e \times a$
e_5	(v_4, v_5)	(e, b)	$e \times b$
e_6	(v_5, v_6)	(e, c)	$e \times c$
e_7	(v_1, v_6)	(c, a)	$c \times a$
e_8	(v_2, v_5)	(b, c)	$b \times c$
e_9	(v_3, v_4)	(a, b)	$a \times b$

Table 5.5 Edge Table of $(((T_1 \cap H) - T_2) - T_3) \cap C$

5.7.4 Edge-Identification Information

The constructions in the previous section demonstrate that we must specify the following information to identify edge segments on space curves unambiguously:

1. The geometry of the carrier; for example, as the intersection of two surfaces
2. The bounding vertices of the edge
3. The intended curve branch
4. The orientation of the branch, at each vertex

The branch orientation and identification are needed because a vertex of the edge may be at a curve singularity.

Let (f, g) denote the intersection of the surfaces f and g, and consider a singular point p on it. At such a point, we have one or more distinct branches of the curve. In general, it is not possible to isolate one of the branches by selecting a better choice of g or by using additional surfaces to intersect with. If singular points are confined to vertices, then the edge segment is homeomorphic to a line, and therefore can be identified unambiguously by an interior point of the edge, as shown in Figure 5.14. The strength of this method is its simplicity. A drawback is its inconvenience: To locate the correct branch and direction at each vertex, we must trace the edge, beginning with the interior point, in both directions.

Another way to identify the branch is to give directional information at the vertices. In the simplest case, the desired branch is identified by the di-

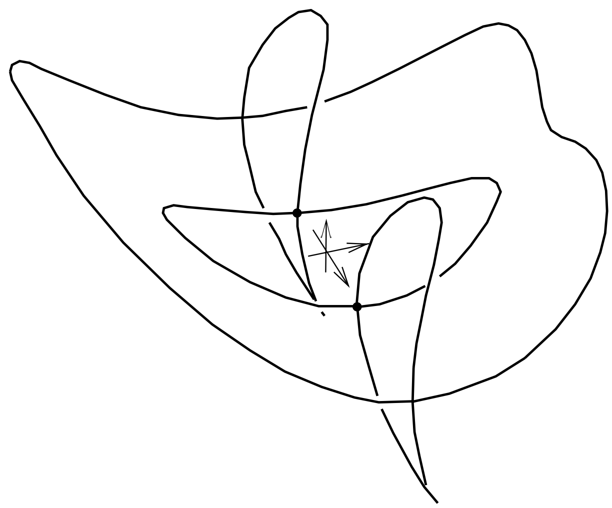

Figure 5.12 Complete Torus/Cylinder Intersection

rected tangent at each vertex. This method suffices for all *nodal* singularities; that is, for singularities at which locally the curve consists of a number of continuous branches that intersect transversally. More difficult singularities are shown in Figures 5.15 and 5.16. The singularity in Figure 5.15 is a *cusp*. The curve has only one branch at the point and the branch is singular. The

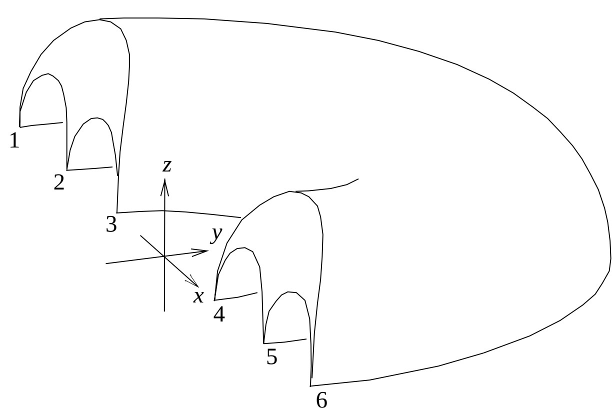

Figure 5.13 Second Interpretation of the Representation

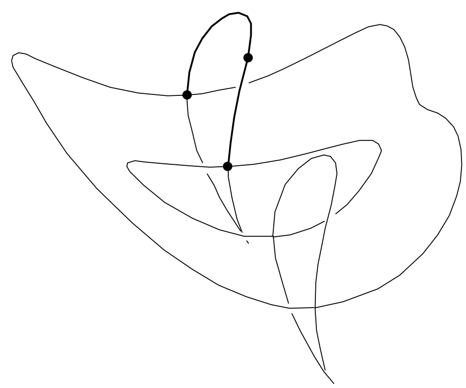

Figure 5.14 Branch Identification by Interior Point

singularity in Figure 5.16 is a *tacnode*. Here, the curve has two branches that intersect tangentially. Intuitively, such singularities require information about higher-order derivatives. This information may be based on the quadratic transformations explained in Section 6.5.2 of Chapter 6. Note that the singularities shown in Figures 5.15 and 5.16 arise for surfaces of fairly low degree: The cusp is the intersection of the cubic surface $y^2 - x^3 + z = 0$ with the plane $z = 0$, and the tacnode is the intersection of the parabolic cylinder $z - y^2 = 0$ with the quartic surface $z - x^4 - y^4 = 0$, approximately a figure of revolution.

The ambiguities constructed here depend on the fact that the curves and surfaces involved are not parametric. Indeed, when edges and faces can be defined in terms of domains on parametric curves and surfaces, the informa-

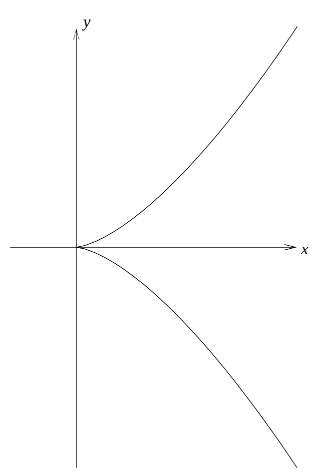

Figure 5.15 Cuspidal Singularity of $y^2 - x^3 + z = 0 \cap z = 0$

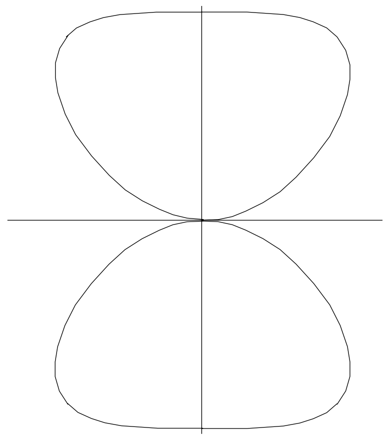

Figure 5.16 Tacnodal Singularity of $z - x^4 - y^4 = 0 \cap z - y^2 = 0$

tion listed here is easily derived from the parameterization and the domain. Unfortunately, many space curves are not parameterizable — even the intersection curves of parameterizable surfaces. For example, the intersection of two cylinders is, in general, a space curve that cannot be represented in a rational parametric form.

5.8 Notes and References

Many books on affine and projective spaces explain the details sketched in Section 5.2. See, for example, Klein (1925), or Semple and Kneebone (1952).

The sections on implicit and parametric representations present standard material from algebraic geometry. Brieskorn and Knörrer (1986) contains an exposition of many of these concepts, and illustrates them with many examples. More condensed and rigorous accounts can be found in van der Waerden (1939), Walker (1950), and many other books on algebraic curves and algebraic geometry. The following chapters present additional information on the intuition behind the algebraic concepts.

The geometry of conics has been studied in great detail by many authors. A nice exposition is found in Hilbert and Cohn-Vossen (1952). Formulae for determining the type of conics and quadrics directly from the coefficients of their equations, and methods for transforming conics to standard form through translation and rotation, can be found in mathematical handbooks, including Bronstein and Semendjajew (1961).

A good entry into the literature on special classes of parametric surfaces is the survey article by Böhm, Farin, and Kahmann (1984). Textbooks on the subject include Mortenson (1985), Bartels, Beatty, and Barsky (1987), and Farin (1988).

The first method for parameterizing conics and quadrics uses a method due to Jacobi. Our exposition is adapted from Golub and van Loan (1983). Jacobi rotations are numerically stable and the iteration has quadratic convergence. The second method for parameterizing conics and quadric surfaces is from Abhyankar and Bajaj (1987a).

Abhyankar and Bajaj (1987b) give the method for parameterizing cubic curves. Their paper also gives a technique for parameterizing cubic surfaces, but the resulting parameterization is not necessarily faithful. A different method for parameterizing cubic surfaces is given in Sederberg and Snively (1987).

Abhyankar and Bajaj (1987c,d) present techniques for parameterizing general algebraic curves. By extension, the method for parameterizing plane algebraic curves also determines whether the curve has genus zero.

Monoids were well known in the nineteenth century, and were used as a tool to classify algebraic space curves. In the more recent literature, they are presented again by Sederberg (1983), where they are called *dual forms* in recognition of the ease of converting between the parametric and the implicit form.

Classical elimination theory has developed a number of different resultant formulations, including formulations that achieve the elimination of two variables in a single step. An early systematic exposition can be found in Netto (1892). Technically, the resultant formulates a system of linear equations symbolically. Macaulay (1902 and 1916) recognized that the extraneous factors are related to the presence of dependent equations, and attempts to eliminate these factors by identifying them as suitable minors of a larger determinant.

Recent interest in elimination theory was stimulated by Sederberg's thesis (1983), which explains the resultant formulations of Sylvester, Bezout, and Dixon. Sederberg (1983) advocates the utility of resultants for implicitizing parametric curves and surfaces. However, as noted in Sederberg and Parry (1986a), using resultants becomes unattractive for curves of degree higher than 4. Many symbolic algebra systems such as Macsyma have implemented the Sylvester resultant.

Macaulay's idea of eliminating dependencies by dividing by a minor has been pursued algorithmically in Canny (1986). The conversion from parametric to implicit form can be done directly by formulating a linear system numerically rather than symbolically. In that case, no extraneous factors are present. See also Chuang and Hoffmann (1989).

The material on boundary-representation ambiguities is adapted from Hoffmann and Hopcroft (1987c). The proposal to identify edge segments by an interior, regular curve point is due to Requicha (1980b).