Chapter 6

Surface Intersections

Evaluating the intersection of two surfaces is a recurring operation in solid modeling — for example, when intersecting B-rep objects. Surface intersection is not an easy problem, and continues to be an active topic of research. Some of the reasons for this continued activity are not hard to identify: A good surface-intersection technique has to balance three conflicting goals: *efficiency*, *robustness*, and *accuracy*.

Typically, a numerical algorithm is efficient, but is not fully robust and so may fail in certain cases. Furthermore, the accuracy a numerical method can deliver varies with the surface degree, with the local surface geometry at the intersection curve, and with the angle at which the surfaces intersect. Algorithms based on exact arithmetic, on the other hand, are fully robust and accurate, but are normally slow. Perhaps the goals of efficiency, robustness, and accuracy cannot be met simultaneously without some compromises, and we might have to negotiate those compromises judiciously, as appropriate for the particular application. Further research is needed to clarify this picture.

In this chapter, we will look at tracing approaches to evaluating surface intersection. Surface intersections can be traced directly, or we can reformulate the intersection problem such that other curves are traced from which information about the surface intersection is computed. In the purest version, a curve-tracing scheme performs the following conceptual operation repeatedly:

At a point p on the intersection, a local approximation of the curve is constructed; for example, the curve tangent at p. By stepping along the approximation a specific distance, we obtain an estimate of a next curve point that we then refine using an iterative method.

Such an algorithm requires solving the following subproblems:

1. Find an initial point p on the intersection curve.
2. Determine a local approximant at p.
3. Select a suitable step size and step along the approximant.
4. Refine the new point estimate to a curve point.

We will not consider how to find an initial starting point, but will concentrate on the remaining steps.

Tracing schemes can be augmented and generalized. We consider two major ways to do this:

- Certain surface operations are naturally formulated using several algebraic equations in more than three variables. This motivates extending the numerical tracing schemes to work in n-dimensional spaces, where $n > 3$.
- Surface intersections can always be mapped to the equivalent problem of evaluating a plane algebraic curve. An attractive aspect of this approach is its ability to cope with singular curve points, a traditional weakness of numerical curve-tracing algorithms.

6.1 Chapter Overview

First, we explain a purely numerical tracing method for evaluating the intersection of two implicit surfaces, $f(x, y, z) = 0$ and $g(x, y, z) = 0$. Technically, the approximant used at a current curve point p is a truncated Taylor expansion of the intersection. The step length is determined adaptively, and the next point estimate is refined iteratively using the Newton–Raphson method. This method is efficient and, when implemented carefully, accurate. However, it is not fully robust and will fail in areas where the intersection curve is singular or nearly so.

The Taylor approximant used in the numerical method is derived by solving a certain system of linear equations whose coefficients depend on the partial derivatives of the two surfaces. It turns out that this formulation is a special case of solving a system of $n-1$ algebraic equations in n variables, assuming that the equations are independent and the corresponding hypersurfaces intersect transversally. This observation can be applied in different ways. For example, we can formulate the intersection of two parametric surfaces equivalently as solving three algebraic equations in four variables.

Other applications include intersecting derived surfaces, including offset surfaces. We explore such applications in Section 6.3. An advantage of the higher-dimensional formulation is that the algebraic degrees involved are often low, and this appears to increase the numerical accuracy of the method in many cases. A disadvantage is that the processing time at each point is greater, since the linear system solved to find the curve approximant is now of size $(n-1) \times n$, where n is the number of variables.

The intersection of two algebraic surfaces is an algebraic space curve, whether the surfaces have been specified implicitly or in parametric form. It is known that every algebraic space curve can be mapped to a plane algebraic curve. In consequence, surface intersection can be approached as follows:

1. Map the surface intersection to a plane algebraic curve $f(u, v) = 0$.
2. Evaluate the plane curve f.
3. Map the points of f back to points on the surface intersection.

In Section 6.4, we present a number of techniques for mapping surface intersections to plane algebraic curves. Using these mapping techniques, it is therefore possible to trace surface intersections by equivalently tracing plane curves. In Section 6.5, we describe a method for evaluating a plane algebraic curve that is capable of dealing with curve singularities. The idea of the method is as follows. Trace f with an ordinary numerical method. When approaching a singularity at a curve point p, apply a transformation that locally changes f to another curve g that is not singular at the corresponding point. Trace the transformed curve g and map each point of g back to f. After the singularity of f has been passed, resume tracing f. The method integrates numerical and symbolic computation. An advantage is its ability to cope with curve singularities. A disadvantage is that the map from a surface intersection to a plane algebraic curve can be difficult to construct.

Tracing a surface intersection in n-dimensional space and mapping the problem to a plane algebraic curve, are, in a sense, two extremes. Generally, when given $n-1$ equations in n variables, we have the option of eliminating none, some, or all but two of the variables. This implies that there are tradeoffs that should be explored. Such tradeoffs are not yet well understood.

6.2 Intersecting Two Implicit Surfaces Numerically

Given two implicit surfaces, $f(x, y, z) = 0$ and $g(x, y, z) = 0$, and a point $p = (p_x, p_y, p_z)$ on their intersection, we wish to trace the intersection curve, beginning at p. The tracing direction is fixed by the cross-product convention explained in the previous chapter. So, a *positive* trace proceeds in the direction $\nabla f \times \nabla g$, whereas a *negative* trace proceeds in the direction $-\nabla f \times \nabla g$.

We assume that, at each point p, the surface gradients ∇f and ∇g are linearly independent; that is, the two surfaces intersect transversally at p. In that case, the intersection curve is regular at p. If the gradients vanish or are linearly dependent, then the curve has a singularity at p and the numerical approach cannot be used without considerable changes and additions.

Assume that p is a regular point on the intersection of f and g. There is a neighborhood of p in which a local parameterization of the intersection curve exists. The local parameterization is a vector-valued function

$$\mathbf{r}(s) = \begin{pmatrix} r_x(s) \\ r_y(s) \\ r_z(s) \end{pmatrix}$$

of a scalar variable s. Note that $r_x(s)$ denotes the x component of the vector $\mathbf{r}(s)$. Analogously, $r_y(s)$ and $r_z(s)$ denote the y and z component, respectively.

It can be shown that the function $\mathbf{r}(s)$ is analytic in a neighborhood of p. By Taylor's theorem, therefore, this function may be written

$$\mathbf{r}(s) = \mathbf{r}(0) + s\mathbf{r}'(0) + \frac{s^2}{2}\mathbf{r}''(0) + \frac{s^3}{6}\mathbf{r}'''(0) + \cdots$$

where $p = \mathbf{r}(0)$. The value of the first derivative of $\mathbf{r}(s)$ at $s = 0$ is $\mathbf{r}'(0)$, that of the second derivative is $\mathbf{r}''(0)$, and so on. The tracing procedure repeats the following steps:

1. At the curve point p, construct a local approximant of $\mathbf{r}(s)$, to some order.[1]

2. Using this approximant and a step value s_0, determine the next point $q = \mathbf{r}(s_0)$.

3. By Newton iteration, bring q closer to the intersection of f and g.

Typically, the order of approximation is fixed. First-order approximations use the curve tangent at p as the local approximant. This is often implemented because the tangent is so easy to compute, as $\mathbf{t} = \nabla f \times \nabla g$. Higher-order approximants allow larger steps. There is a tradeoff between the added time needed to compute a higher-order approximant, and the time saved by the ability to take larger steps. Degree-3 approximants seem to provide a good balance in that the determination of the approximant is not too costly, and the approximant accounts for both the curvature and the torsion at p.

[1] In the following discussion, $\mathbf{r}(s)$ will also denote the approximant.

6.2.1 Construction of the Approximant

We view the intersection of f and g as a vector function $\mathbf{r}(s)$, parameterized by the scalar variable s, with $p = \mathbf{r}(0)$, and write

$$\mathbf{r}(s) = \mathbf{r}(0) + s\mathbf{r}'(0) + \frac{s^2}{2}\mathbf{r}''(0) + \frac{s^3}{6}\mathbf{r}'''(0) + \cdots$$

since the curve is not singular at p by assumption. The approximant is an initial segment of this series and is determined by finding the components of the derivatives of $\mathbf{r}$, up to some order. Technically, this involves the formulation of a linear system, which always has the following structure:

$$\begin{aligned} \nabla f \cdot \mathbf{r}^{(m)}(0) &= b_{f,m} \\ \nabla g \cdot \mathbf{r}^{(m)}(0) &= b_{g,m} \end{aligned} \tag{6.1}$$

Here, $\mathbf{r}^{(m)}(s)$ denotes the m^{th} derivative of $\mathbf{r}(s)$ by s. The coefficients $b_{f,m}$ and $b_{g,m}$ depend on the partial derivatives of f and g at p, and the derivatives of $\mathbf{r}$ up to order $m - 1$.

Since the system is underdetermined, it does not have a unique solution, and we must make certain choices. These choices have a geometric interpretation, and will result in an approximant where the values for $\mathbf{r}''$ and $\mathbf{r}'''$ are explicitly related to curvature and torsion at p.

Setting Up the Linear System

We determine the derivative values, $\mathbf{r}'(0)$, $\mathbf{r}''(0)$, $\mathbf{r}'''(0)$, from the partial derivatives of f and of g. When $p = (p_x, p_y, p_z)$ is a regular point of the surface f, by Taylor's theorem, there exists a neighborhood of p in that

$$f(x, y, z) = f(p_x + \delta_x, p_y + \delta_y, p_z + \delta_z) = \sum_{i,j,k} f_{i,j,k}\delta_x^i \delta_y^j \delta_z^k$$

for real numbers δ_x, δ_y, and δ_z. The coefficients $f_{i,j,k}$ in the sum denote expressions

$$f_{i,j,k} = \frac{1}{i!j!k!} \frac{\partial^{i+j+k}}{\partial x^i \partial y^j \partial z^k} f(p_x, p_y, p_z)$$

Let $f(p_x, p_y, p_z)$ and $f(p_x + \delta_x, p_y + \delta_y, p_z + \delta_z)$ be points on the curve $\mathbf{r}(s)$. Assuming that $p = \mathbf{r}(0)$ and $(p_x + \delta_x, p_y + \delta_y, p_z + \delta_z) = \mathbf{r}(s)$, we set

$$\begin{aligned} \delta_x &= \mathbf{r}'_x s + \mathbf{r}''_x s^2/2 + \mathbf{r}'''_x s^3/6 + \cdots \\ \delta_y &= \mathbf{r}'_y s + \mathbf{r}''_y s^2/2 + \mathbf{r}'''_y s^3/6 + \cdots \\ \delta_z &= \mathbf{r}'_z s + \mathbf{r}''_z s^2/2 + \mathbf{r}'''_z s^3/6 + \cdots \end{aligned}$$

where $\mathbf{r}'_x$ denotes the x-component of the vector $\mathbf{r}'$, and so on.[2] Then, we have

$$\begin{aligned}
(\delta_x)^2 &= (\mathbf{r}'_x)^2 s^2 + \mathbf{r}'_x \mathbf{r}''_x s^3 + \cdots \\
(\delta_x)^3 &= (\mathbf{r}'_x)^3 s^3 + \cdots \\
\delta_x \delta_y &= \mathbf{r}'_x \mathbf{r}'_y s^2 + (\mathbf{r}''_x \mathbf{r}'_y + \mathbf{r}'_x \mathbf{r}''_y) s^3/2 + \cdots \\
\delta_x \delta_y \delta_z &= \mathbf{r}'_x \mathbf{r}'_y \mathbf{r}'_z s^3 + \cdots
\end{aligned}$$

and so on.

Since the curve $\mathbf{r}$ is on f, substitution of these quantities must yield identically zero; hence, the coefficient of s^m must vanish for each m. For $m = 1$, 2, 3, we therefore obtain the equations

$$\begin{aligned}
f_{1,0,0}\mathbf{r}'_x + f_{0,1,0}\mathbf{r}'_y + f_{0,0,1}\mathbf{r}'_z &= b_{f,1} \\
f_{1,0,0}\mathbf{r}''_x + f_{0,1,0}\mathbf{r}''_y + f_{0,0,1}\mathbf{r}''_z &= b_{f,2} \\
f_{1,0,0}\mathbf{r}'''_x + f_{0,1,0}\mathbf{r}'''_y + f_{0,0,1}\mathbf{r}'''_z &= b_{f,3}
\end{aligned}$$

The righthand sides $b_{f,k}$ are computed from the partials of f and lower-order derivatives of $\mathbf{r}$

$$b_{f,1} = 0$$

$$\begin{aligned}
b_{f,2} = -2[&f_{2,0,0}(\mathbf{r}'_x)^2 + f_{0,2,0}(\mathbf{r}'_y)^2 + f_{0,0,2}(\mathbf{r}'_z)^2 \\
&+ f_{1,1,0}\mathbf{r}'_x\mathbf{r}'_y + f_{1,0,1}\mathbf{r}'_x\mathbf{r}'_z + f_{0,1,1}\mathbf{r}'_y\mathbf{r}'_z]
\end{aligned}$$

$$\begin{aligned}
b_{f,3} = -6[&f_{2,0,0}\mathbf{r}'_x\mathbf{r}''_x + f_{0,2,0}\mathbf{r}'_y\mathbf{r}''_y + f_{0,0,2}\mathbf{r}'_z\mathbf{r}''_z + f_{1,1,0}(\mathbf{r}''_x\mathbf{r}'_y + \mathbf{r}'_x\mathbf{r}''_y)/2 \\
&+ f_{1,0,1}(\mathbf{r}''_x\mathbf{r}'_z + \mathbf{r}'_x\mathbf{r}''_z)/2 + f_{0,1,1}(\mathbf{r}''_y\mathbf{r}'_z + \mathbf{r}'_y\mathbf{r}''_z)/2 \\
&+ f_{3,0,0}(\mathbf{r}'_x)^3 + f_{0,3,0}(\mathbf{r}'_y)^3 + f_{0,0,3}(\mathbf{r}'_z)^3 \\
&+ f_{2,1,0}(\mathbf{r}'_x)^2\mathbf{r}'_y + f_{1,2,0}\mathbf{r}'_x(\mathbf{r}'_y)^2 + f_{2,0,1}(\mathbf{r}'_x)^2\mathbf{r}'_z \\
&+ f_{1,0,2}\mathbf{r}'_x(\mathbf{r}'_z)^2 + f_{0,2,1}(\mathbf{r}'_y)^2\mathbf{r}'_z + f_{0,1,2}\mathbf{r}'_y(\mathbf{r}'_z)^2 + f_{1,1,1}\mathbf{r}'_x\mathbf{r}'_y\mathbf{r}'_z]
\end{aligned}$$

With the $b_{f,m}$ as the right-hand sides, we can rewrite these equations in vectorial notation as

$$\begin{aligned}
\nabla f \cdot \mathbf{r}' &= b_{f,1} \\
\nabla f \cdot \mathbf{r}'' &= b_{f,2} \\
\nabla f \cdot \mathbf{r}''' &= b_{f,3}
\end{aligned}$$

---

[2] We write $\mathbf{r}$ instead of $\mathbf{r}(0)$, $\mathbf{r}'$ instead of $\mathbf{r}'(0)$, and so on.

The equations for g are developed analogously. In particular, we have

$$\begin{aligned}\nabla g \cdot \mathbf{r}' &= b_{g,1}\\ \nabla g \cdot \mathbf{r}'' &= b_{g,2}\\ \nabla g \cdot \mathbf{r}''' &= b_{g,3}\end{aligned}$$

where

$$\begin{aligned}
b_{g,1} &= 0\\
b_{g,2} &= -2[g_{2,0,0}(\mathbf{r}'_x)^2 + g_{0,2,0}(\mathbf{r}'_y)^2 + g_{0,0,2}(\mathbf{r}'_z)^2\\
&\quad + g_{1,1,0}\mathbf{r}'_x\mathbf{r}'_y + g_{1,0,1}\mathbf{r}'_x\mathbf{r}'_z + g_{0,1,1}\mathbf{r}'_y\mathbf{r}'_z]\\
b_{g,3} &= -6[g_{2,0,0}\mathbf{r}'_x\mathbf{r}''_x + g_{0,2,0}\mathbf{r}'_y\mathbf{r}''_y + g_{0,0,2}\mathbf{r}'_z\mathbf{r}''_z + g_{1,1,0}(\mathbf{r}''_x\mathbf{r}'_y + \mathbf{r}'_x\mathbf{r}''_y)/2\\
&\quad + g_{1,0,1}(\mathbf{r}''_x\mathbf{r}'_z + \mathbf{r}'_x\mathbf{r}''_z)/2 + g_{0,1,1}(\mathbf{r}''_y\mathbf{r}'_z + \mathbf{r}'_y\mathbf{r}''_z)/2\\
&\quad + g_{3,0,0}(\mathbf{r}'_x)^3 + g_{0,3,0}(\mathbf{r}'_y)^3 + g_{0,0,3}(\mathbf{r}'_z)^3\\
&\quad + g_{2,1,0}(\mathbf{r}'_x)^2\mathbf{r}'_y + g_{1,2,0}\mathbf{r}'_x(\mathbf{r}'_y)^2 + g_{2,0,1}(\mathbf{r}'_x)^2\mathbf{r}'_z\\
&\quad + g_{1,0,2}\mathbf{r}'_x(\mathbf{r}'_z)^2 + g_{0,2,1}(\mathbf{r}'_y)^2\mathbf{r}'_z + g_{0,1,2}\mathbf{r}'_y(\mathbf{r}'_z)^2 + g_{1,1,1}\mathbf{r}'_x\mathbf{r}'_y\mathbf{r}'_z]
\end{aligned}$$

We put these equations into matrix form. A is a 2×3 matrix whose rows are the gradients of f and of g, and all partials of f and of g are evaluated at p. Then the system is

$$A\mathbf{r}^{(m)} = \begin{pmatrix} b_{f,m} \\ b_{g,m} \end{pmatrix}$$

Although this system is only 2×3, solving it without giving proper attention to its numerical properties will waste accuracy in the solution. Hence, we should carefully choose a numerically stable solution technique. A good choice is *singular value decomposition*, sketched in Section 6.1.4.

As output, singular value decomposition delivers, in our case, two scalars, σ_1 and σ_2; three orthonormal vectors in three-dimensional space, U_1, U_2, U_3; and an orthogonal 2×2 matrix V. From these quantities, we construct a solution of the form

$$\mathbf{r}^{(m)} = \alpha_m U_1 + \beta_m U_2 + \gamma_m U_3$$

where the coefficients α_m and β_m are determined by V, σ_1, and σ_2. The details are deferred to Section 6.2.4.

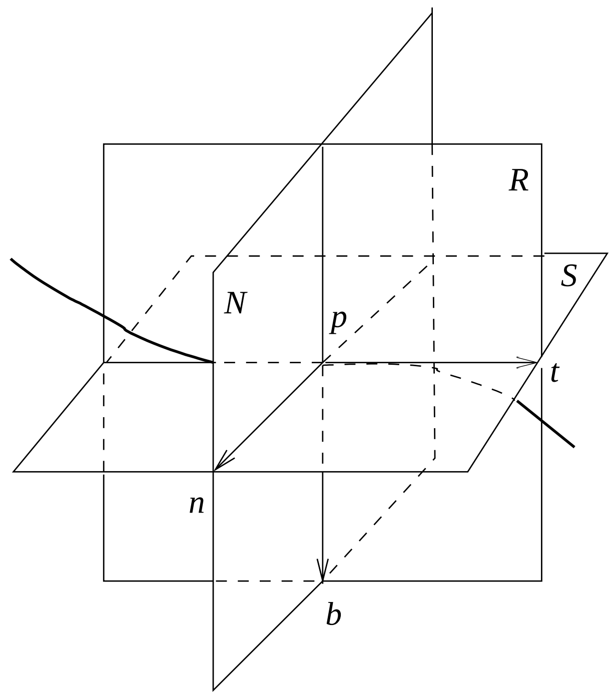


Figure 6.1 The Space-Curve Triad

Choosing the Undetermined Coefficients

The linear system is underdetermined and has an infinity of solutions. So, choices must be made for the γ_m to arrive at a canonical solution. Our strategy is to choose values such that the derivative values reveal some of the intrinsic geometric structure of the curve at the point p.

From differential geometry, we recall that at the point p of a space curve, the *moving triad* forms a natural local coordinate system.[3] The triad consists of three orthonormal vectors, the *tangent vector* $\mathbf{t}$, the *principal normal vector* $\mathbf{n}$, and the *binormal vector* $\mathbf{b}$, where $\mathbf{b} = \mathbf{t} \times \mathbf{n}$. Their directions are defined by the tangent, the curvature, and the torsion of the space curve.

The curve *tangent* t at the point p is the limiting position of curve secants (p, q), where q approaches p. The plane perpendicular to t is the *normal plane* N at p. We consider a plane through the tangent and an additional curve point r. As r approaches p, this plane approaches as limit position the *osculating plane* S. The perpendicular to t in the osculating plane is the *principal normal* n. The plane perpendicular to both N and S is the *rectifying plane* R. The perpendicular to t in the rectifying plane is the *binormal* b; see also Figure 6.1. At a regular point, the curve intersects the osculating plane but remains on one side of the rectifying plane.

We consider three points r, p, and q on the curve. They define a circle. As r and q approach p, the limit position of the circle is the *circle of curvature*

[3]The moving triad is also called the *Frenet frame*.

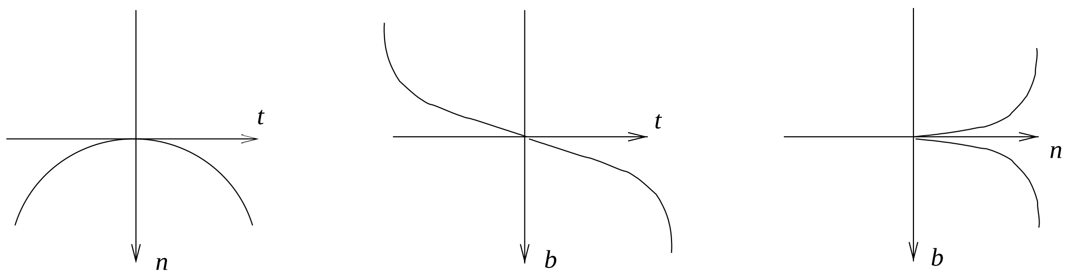


Figure 6.2 Local Curve Projections

.

and lies in S. Its radius ρ is the *radius of curvature* at p. Now consider four curve points. If the points are not coplanar, they define a sphere. As three of the points approach the fourth point p, the limiting position of this sphere is the *osculating sphere*. The center of the osculating sphere is in the normal plane, and the sphere intersects the osculating plane in the circle of curvature.

Intuitively, the torsion at p is obtained by considering how the osculating plane changes with p. Consider the angle between the osculating planes at p and a nearby curve point q. As q approaches p, the ratio between this angle and the arc length (p, q) approaches as limit the torsion T of the curve at p.

We orient the curve at p by choosing a direction for the tangent t, and we denote the unit vector in this direction by $\mathbf{t}$. Then, we orient the normal toward the concave curve side; that is, toward the center of the circle of curvature. The unit vector in this direction will be $\mathbf{n}$. Finally, we orient the binormal by the vector $\mathbf{b} = \mathbf{t} \times \mathbf{n}$, as shown in Figure 6.1. With these conventions of orientation, the projection of the space curve onto each of the three planes is locally as shown in Figure 6.2.

As the point p moves on the space curve the vectors $\mathbf{t}$, $\mathbf{n}$, and $\mathbf{b}$ vary obeying the *Frenet–Serret formulae.*

$$\frac{d\mathbf{t}}{ds} = \kappa\mathbf{n} \qquad \frac{d\mathbf{b}}{ds} = -T\mathbf{n} \qquad \frac{d\mathbf{n}}{ds} = T\mathbf{b} - \kappa\mathbf{t} \tag{6.2}$$

Here, s is the arc length, $\kappa = 1/\rho$ is the curvature, and T is the torsion of the curve.

Now U_3, as determined by singular value decomposition, is a unit vector in the tangent direction. So, we choose $\gamma_1 = \pm 1$, depending on the tracing direction, and observe that $\mathbf{r}' = \pm U_3$. By choosing γ_2 and γ_3 properly, we relate the higher-order derivatives $\mathbf{r}''$ and $\mathbf{r}'''$ to the moving triad at p. Since

$$\mathbf{r}'' = \frac{d\mathbf{t}}{ds} = \kappa\mathbf{n}$$

$\mathbf{r}''$ should be perpendicular to $\mathbf{t}$. So, we choose $\gamma_2 = 0$. Hence, the curvature at p will be

$$\kappa = \sqrt{\alpha_2^2 + \beta_2^2}$$

Finally, we have

$$\mathbf{r}'''(s) = \frac{d}{ds}(\kappa\mathbf{n}) = \frac{d\kappa}{ds}\mathbf{n} + \kappa\frac{d\mathbf{n}}{ds} = \kappa'\mathbf{n} + \kappa T\mathbf{b} - \kappa^2\mathbf{t}$$

But $\mathbf{r}''' = \alpha_3 U_1 + \beta_3 U_2 + \gamma_3\mathbf{t}$, so, by orthogonality, we have $\gamma_3 = -\kappa^2$. Moreover, we can compute the torsion at p by projecting $\mathbf{r}'''$ onto $\mathbf{b}$ and dividing the length of the projected vector by κ.

Recall that the method for determining the quantities γ_1, γ_2, and γ_3 just presented is not the only one. We can understand different strategies by interpreting them geometrically. For simplicity, we consider first-order approximants and interpret the effect of choosing a value for γ_1. Now, with $\gamma_1 = 1$, the estimated curve point obtained as

$$q = \mathbf{r}(0) + s\,\mathbf{r}'(0)$$

has distance s from p. If $0 < \gamma_1 < 1$, then the distance is less than s; with $\gamma_1 > 1$, it is greater than s. Hence, the choices for γ_m determine how $\mathbf{r}(s)$ is parameterized.

6.2.2 Selection of Step Size

We have constructed an approximant $\mathbf{r}(s)$ to the curve at p; now we must choose a step length s_0 to obtain a subsequent curve-point estimate $\mathbf{r}(s_0)$. Choosing a safe step length requires understanding the radius of convergence of the full Taylor series. To this end, we consider each coordinate of $\mathbf{r}$ separately as a function of s. For each coordinate of $\mathbf{r}(s)$, we have a function

$$F(s) = \sum_{n=0}^{\infty} a_n s^n \tag{6.3}$$

This series will converge absolutely for all values of s that satisfy $|s| < R$, and will diverge for all values $|s| \geq R$, where

$$R = \lim \sup |a_n|^{1/n}$$

We assume that $R = 0$ whenever $|a_n|^{1/n}$ is unbounded. Therefore, the radius of convergence for the full Taylor series of $\mathbf{r}$ is the minimum R of the three coordinate functions $F(s)$.

In practice, the determination of R is difficult except in those cases where simple recurrences or closed-form expressions can be given for the coefficients a_n. Thus, we opt for a simpler heuristic in which the contribution of the quadratic and cubic terms to the next point estimate is kept small. For example, since $\mathbf{r}'$ has unit length, we may choose s_0 such that both

$$\frac{\|s_0^2 \mathbf{r}''(0)\|}{2} < \frac{|s_0|}{10} \qquad \text{and} \qquad \frac{\|s_0^3 \mathbf{r}'''(0)\|}{6} < \frac{|s_0|}{10}$$

Since the step sizes could become arbitrarily small, a minimum step size should also be specified. This simple strategy does well in many cases.

6.2.3 Newton Iteration

At the point p, we have constructed a third-order approximant $\mathbf{r}(s)$, we have determined adaptively a step length s_0, and now we have a new point estimate $q = \mathbf{r}(s_0)$. Using Newton iteration, we refine this estimate until we are on the intersection of f and g with acceptable accuracy. The iteration is based on the following, first-order approximation of the two surfaces:

$$\begin{aligned} \nabla f(q_k) \cdot \Delta_k &= -f(q_k) \\ \nabla g(q_k) \cdot \Delta_k &= -g(q_k) \end{aligned}$$

where $\Delta_k = (\delta_x, \delta_y, \delta_z)^T$. Note that this system has the same structure as does system (6.1). Solving it for Δ_k, we obtain the next point estimate as

$$q_{k+1} = q_k + \Delta_k$$

As in the approximant construction, we solve the linear system using singular value decomposition. For the solution Δ_k, we set the coefficient of U_3 to zero, since it represents lateral movement that will not improve the quality of the new estimate significantly. We continue with the iteration until

$$\|q_{k+1} - q_k\| < 10^{-t}\|q_k\|$$

where t is a precision parameter. Typically, we have $t = 10$ for double-precision floating-point computations, and we require two or three iterations to achieve this accuracy.

6.2.4 Singular Value Decomposition

Singular value decomposition is a method for solving systems of linear equations that may be singular. It is a numerically stable method of considerable flexibility, and is a part of many widely available software libraries.

Using Singular Value Decomposition

Assume that we are given the linear system

$$A\mathbf{r} = \mathbf{b} \tag{6.4}$$

From the system matrix A, the singular-value-decomposition algorithm constructs three matrices U, S, and V such that

$$A = VSU^T \tag{6.5}$$

where the matrices have the following properties:

1. The matrices U and V are orthogonal; that is, UU^T and VV^T are the identity matrices.

2. The matrix S is diagonal and its diagonal entries are nonnegative and decreasing.

When we are using the method for the intersection of two implicit surfaces, the matrix V is 2×2, U is 3×3, and S is

$$S = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix}$$

where $\sigma_1 \geq \sigma_2 \geq 0$. If the surfaces intersect transversally, moreover, then $\sigma_2 > 0$. Using equation (6.5), we transform system (6.4) to

$$S(U^T\mathbf{r}) = V^T\mathbf{b}$$

Let U_i be the i^{th} column in U, and denote by $(V^T\mathbf{b})_i$ the i^{th} component of the vector $V^T\mathbf{b}$. Since U is orthogonal, the vectors U_i are also orthogonal. Then it can be proved that the solution of system (6.4) is

$$\mathbf{r} = \alpha U_1 + \beta U_2 + \gamma U_3$$

where the scalar coefficients α and β are given by

$$\begin{aligned} \alpha &= (V^T\mathbf{b})_1/\sigma_1 \\ \beta &= (V^T\mathbf{b})_2/\sigma_2 \end{aligned}$$

The last coefficient γ is arbitrary, and the column vector U_3 is in the null space of A. When intersecting two implicit surfaces, the null space at a regular curve point is spanned by $\nabla f \times \nabla g$. Hence, U_3 is the tangent direction to the curve at $p = \mathbf{r}(0)$. Note, however, that the direction of U_3 could be equal to or opposite the direction $\nabla f \times \nabla g$.

First Phase of the Algorithm

The singular-value-decomposition algorithm transforms the matrix A in two phases. Let A be $m \times n$, where $m \leq n$. In the first phase, A is changed to a matrix B in lower bidiagonal form. This is done using *Householder* transformations that multiply A left and right with certain orthogonal matrices.

$$B = \begin{pmatrix} a_{11} & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots \\ a_{21} & a_{22} & 0 & \ldots & 0 & 0 & 0 & \ldots \\ 0 & a_{32} & a_{33} & \ldots & 0 & 0 & 0 & \ldots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \ldots & a_{m-1m-1} & 0 & 0 & \ldots \\ 0 & 0 & 0 & \ldots & a_{mm-1} & a_{mm} & 0 & \ldots \end{pmatrix}$$

At each step, we zero all elements to the right of the i^{th} diagonal element by multiplying A from the right with a matrix U_i, and then zero all elements below the $(i+1, i)$-element by multiplying from the left with V_i.

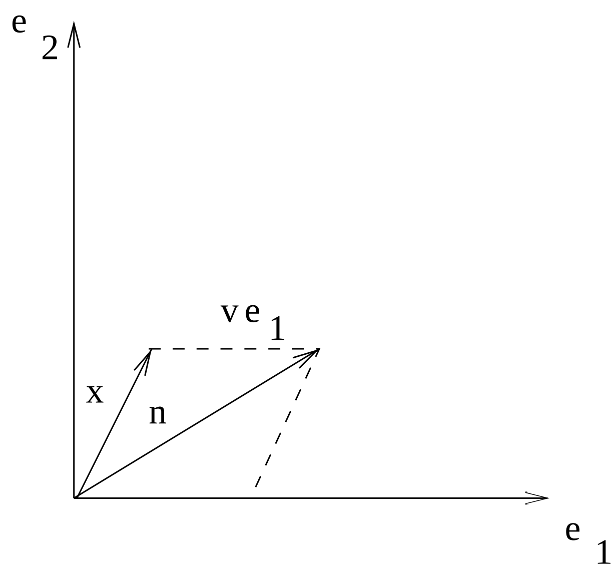


Figure 6.3 The Vector $\mathbf{n}$

A Householder transformation is essentially a reflection about the direction of a column vector $\mathbf{n}$, and is effected by a matrix of the form

$$P = I - \frac{2\mathbf{n}\mathbf{n}^T}{\mathbf{n}^T\mathbf{n}}$$

P is symmetric and orthogonal. Given a column vector $\mathbf{x}$, the vector $\mathbf{n}$ can be chosen such that $P\mathbf{x}$ is a multiple of the unit vector $\mathbf{e}_1 = (1, 0, 0, \ldots)^T$. To do so, we use

$$\mathbf{n} = \mathbf{x} + \nu\mathbf{e}_1$$

where ν is $\|\mathbf{x}\|_2$; see also Figure 6.3. A calculation shows that the matrix P so constructed from $\mathbf{n}$ has the desired effect on $\mathbf{x}$.

We use Householder transformations to zero out blocks of entries in the matrix A. First, we zero all elements to the right of the (i, i)-element. Let $\mathbf{x} = (x_1, \ldots, x_i, x_{i+1}, \ldots, x_n)$ be the i^{th} row in the matrix. We multiply A from the right with an $n \times n$ matrix U_i of the form

$$U_i = \begin{pmatrix} I & 0 \\ 0 & P \end{pmatrix}$$

where I is the $(i-1) \times (i-1)$ identity matrix, and P is the Householder matrix for $(x_i, x_{i+1}, \ldots, x_n)^T$. Next, we zero all elements below the $(i+1, i)$-element by multiplying from the left with a matrix V_i. If $(y_1, ..., y_i, y_{i+1}, \ldots, y_m)^T$ is the i^{th} column of A, then we multiply with

$$V_i = \begin{pmatrix} I & 0 \\ 0 & P \end{pmatrix}$$

where I is $i \times i$ and P is the Householder matrix for $(y_{i+1}, ..., y_n)^T$. In this way, we can change A to a lower bidiagonal form.

Second Phase of the Algorithm

In the second phase of singular value decomposition, an iteration is done to diagonalize the output from the first phase. The iteration uses *Givens* rotations, and it has similarities with the procedure to diagonalize a real symmetric matrix discussed in Chapter 5, Section 5.5.1.

We consider a rotation matrix that differs from the identity only in the four entries (i,i), (i,j), (j,i), (j,j):

$$J(i,j,\theta) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & \vdots & & \vdots & & \\ & \ldots & c & \ldots & -s & \ldots & \\ & & \vdots & \ddots & \vdots & & \\ & \ldots & s & \ldots & c & \ldots & \\ & & \vdots & & \vdots & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

where $s = \sin(\theta)$ and $c = \cos(\theta)$. By a computation analogous to the derivation of Jacobi rotations, we can show that the j^{th} element of the row vector $\mathbf{x} = (..., x_i, ..., x_j, ...)$ is canceled in $\mathbf{x}J(i,j,\theta)$, provided θ is such that

$$\begin{aligned} \cos(\theta) &= x_i/(x_i^2 + x_j^2) \\ \sin(\theta) &= x_j/(x_i^2 + x_j^2) \end{aligned}$$

The resulting vector is equal to $\mathbf{x}$ except for the i^{th} and the j^{th} components.

We apply Givens rotations as follows: In the lower bidiagonal matrix B, we initially zero the element b_{21} with a rotation $J(1,2,\theta)^T B$. Since B is not symmetric, this results in a nonzero entry b_{12}. We cancel the element a_{12} with a rotation $J(1,2,\theta)$. This reintroduces a_{21} and an element a_{31}. The element a_{31} is canceled next in $J(1,3,\theta)^T B$, introducing a nonzero element a_{23}. We

$$\begin{pmatrix} * & & \\ + & * & \\ & * & * \end{pmatrix} \Longrightarrow \begin{pmatrix} * & + & \\ 0 & * & \\ & * & * \end{pmatrix} \Longrightarrow \begin{pmatrix} * & 0 & \\ * & * & \\ + & * & * \end{pmatrix}$$

$$\Longrightarrow \begin{pmatrix} * & & \\ * & * & + \\ 0 & * & * \end{pmatrix} \Longrightarrow \begin{pmatrix} * & & \\ * & * & 0 \\ & * & * \end{pmatrix}$$

Figure 6.4 Sequence of Givens Rotations in Phase 2 of Singular Value Decomposition

handle a_{23} just like we have a_{12}, pushing the unwanted element to position a_{34} with two rotations. In this manner, the unwanted element is percolated down. When it reaches the position $a_{n-1,n}$, it disappears with the next rotation, leaving us again with a matrix in lower bidiagonal form. This sequence is illustrated in Figure 6.4. It can be shown that, when suitably starting out the initial rotation to cancel a_{21}, the magnitude of the off-diagonal entries is diminished with each such sequence of Givens rotations.

Implementation

Our description of the singular-value-decomposition algorithm leaves out a number of important implementation details. First, the structure of the Householder matrices is such that they do not need to be formed explicitly. The vector $\mathbf{n}$ can be used directly, and this results in an $O(n^2)$ cost for each transformation step of A. Similar considerations apply to the second phase. Moreover, we should partition the matrix at each zero off-diagonal entry, treating the resulting blocks of submatrices separately, and permute rows and columns suitably so that the diagonal elements of S are nonincreasing.

6.3 Tracing in Higher Dimensions

The numerical tracing method we have described is not limited to evaluating the intersection of two implicit surfaces in 3-space; but it is easily generalized to intersecting $n-1$ hypersurfaces in n-space, where $n > 3$. We sketch this generalization and discuss several applications.

6.3.1 The Method

We generalize the material of Section 6.2 to the problem of tracing the intersection of a system of $n-1$ algebraic hypersurfaces in n-dimensional space.

Let

$$\begin{aligned} f_1(x_1, x_2, ..., x_n) &= 0 \\ f_2(x_1, x_2, ..., x_n) &= 0 \\ &\vdots \\ f_{n-1}(x_1, x_2, ..., x_n) &= 0 \end{aligned}$$

be the surfaces that must be intersected. We assume that we are given a regular point p on the intersection. The algorithm considers a local parameterization of the intersection, in a neighborhood of p, as a vector function $\mathbf{r}(s)$ of a scalar variable s. Note that $\mathbf{r}(s)$ now has n scalar components, one for each variable x_i. We proceed as follows:

1. At the curve point p, determine an initial segment of the Taylor expansion of $\mathbf{r}(s)$ at p, by solving a linear system of the form

   $$A\mathbf{r}^{(m)} = \mathbf{b}_m$$

   The rows of A are the gradients of the f_i, and the entries of $\mathbf{b}_m$ depend on the partial derivatives of the f_i and the derivatives of $\mathbf{r}$ up to order $m - 1$.

2. Determine a suitable step value s_0, and derive a new curve point estimate $q = \mathbf{r}(s_0)$.

3. Refine the estimate q to a curve point using Newton iteration.

As before, the linear systems are derived by considering the Taylor expansion at p for each surface f_i, and deriving from it expressions for the curve derivatives $\mathbf{r}'$, $\mathbf{r}''$, and so on. Let $f_i^{(k)}$ denote the partial derivative of f_i by x_k, and $f_i^{(k,j)}$ denote the partial derivative of $f_i^{(k)}$ by x_j. Moreover, let $\mathbf{r}_i$ denote the i^{th} component of $\mathbf{r}$. Then, the first derivative of $\mathbf{r}(s)$ is determined from $n - 1$ equations of the form

$$f_i^{(1)}\mathbf{r}_1' + f_i^{(2)}\mathbf{r}_2' + \cdots + f_i^{(n)}\mathbf{r}_n' = 0$$

The second derivative is determined from $n - 1$ equations of the form

$$f_i^{(1)}\mathbf{r}_1'' + f_i^{(2)}\mathbf{r}_2'' + \cdots + f_i^{(n)}\mathbf{r}_n'' = b_{i,2}$$

where the right-hand side is

$$
\begin{array}{rcrcl}
b_{i,2} & = & -(f_i^{(1,1)}\mathbf{r}_1'^{\,2} + f_i^{(2,2)}\mathbf{r}_2'^{\,2} + & \cdots & +f_i^{(n,n)}\mathbf{r}_n'^{\,2}) \\
 & & -2(f_i^{(1,2)}\mathbf{r}_1'\mathbf{r}_2' + f_i^{(1,3)}\mathbf{r}_1'\mathbf{r}_3' + & \cdots & +f_i^{(1,n)}\mathbf{r}_1'\mathbf{r}_n' \\
 & & +f_i^{(2,3)}\mathbf{r}_2'\mathbf{r}_3' + & \cdots & +f_i^{(2,n)}\mathbf{r}_2'\mathbf{r}_n' \\
 & & & \ddots & \\
 & & & & +f_i^{(n-1,n)}\mathbf{r}_{n-1}'\mathbf{r}_n')
\end{array}
$$

The third-order derivative of $\mathbf{r}$ can be obtained analogously.

Using singular value decomposition to solve the system, there are free parameters to be chosen corresponding to the γ_m in the three-dimensional case. Here we choose $\gamma_1 = \pm 1$ and $\gamma_2 = 0$.[4] A suitable stepping length may be determined with the same heuristics as those used in the three-dimensional case. Thereafter, Newton iteration is used to refine the new curve-point estimate. The details are straightforward.

6.3.2 Numerically Intersecting Two Parametric Surfaces

The traditional approach to intersecting two parametric surfaces is to use subdivision and piecewise linear approximation. When an initial intersection point p is known, however, the numerical approach becomes directly applicable, greatly simplifying the problem.

Let the surfaces be given as

$$
\begin{array}{rcl}
x & = & G_{1,1}(u_1, v_1) \\
y & = & G_{1,2}(u_1, v_1) \\
z & = & G_{1,3}(u_1, v_1)
\end{array}
\quad \text{and} \quad
\begin{array}{rcl}
x & = & G_{2,1}(u_2, v_2) \\
y & = & G_{2,2}(u_2, v_2) \\
z & = & G_{2,3}(u_2, v_2)
\end{array}
$$

Then the intersection is given by the equations

$$
\begin{array}{rcccl}
F_1(u_1, v_1, u_2, v_2) & = & G_{1,1}(u_1, v_1) - G_{2,1}(u_2, v_2) & = & 0 \\
F_2(u_1, v_1, u_2, v_2) & = & G_{1,2}(u_1, v_1) - G_{2,2}(u_2, v_2) & = & 0 \\
F_3(u_1, v_1, u_2, v_2) & = & G_{1,3}(u_1, v_1) - G_{2,3}(u_2, v_2) & = & 0
\end{array}
$$

[4] A generalization of the Frenet–Serret formulae to n dimensions exists and can be used to devise a strategy for determining γ_3. Alternatively, we can designate a subset of three coordinates and determine γ_3 based on them alone, using the method of Section 6.2. This alternative is natural in some applications, including offset surface intersection, discussed later.

that is, by three equations in the four unknowns u_1, v_1, u_2, and v_2. These equations define a curve $\mathbf{r}(s)$ in four-dimensional space:

$$\mathbf{r}(s) = \begin{pmatrix} u_1(s) \\ v_1(s) \\ u_2(s) \\ v_2(s) \end{pmatrix}$$

The surface intersection is recovered from $\mathbf{r}(s)$ using the functions $G_{j,k}$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} G_{1,1}(u_1(s), v_1(s)) \\ G_{1,2}(u_1(s), v_1(s)) \\ G_{1,3}(u_1(s), v_1(s)) \end{pmatrix} = \begin{pmatrix} G_{2,1}(u_2(s), v_2(s)) \\ G_{2,2}(u_2(s), v_2(s)) \\ G_{2,3}(u_2(s), v_2(s)) \end{pmatrix}$$

Example 6.1: Consider intersecting two bicubic surfaces. Assume that the first surface is given by

$$\begin{aligned} x &= G_{1,1}(u_1, v_1) &&= 3v_1(v_1 - 1)^2(u_1 - 1)^3 + 3u_1 \\ y &= G_{1,2}(u_1, v_1) &&= 3u_1(u_1 - 1)^2 v_1^3 + 3v_1 \\ z &= G_{1,3}(u_1, v_1) &&= (u_1^2 - 5u_1 + 5)v_1^3 - 3(u_1^3 + 6u_1^2 - 9u_1 + 1)v_1^2 \end{aligned}$$

and the second surface by

$$\begin{aligned} x &= G_{2,1}(u_2, v_2) &&= u_2^3 v_2^2 - u_2^3 \\ y &= G_{2,2}(u_2, v_2) &&= u_2^2 v_2 + 2u_2^3 v_2^3 \\ z &= G_{2,3}(u_2, v_2) &&= u_2 v_2^3 + u_2^2 v_2 \end{aligned}$$

Then the equations to be traced are

$$\begin{aligned} 3v_1(v_1 - 1)^2(u_1 - 1)^3 + 3u_1 - u_2^3 v_2^2 + u_2^3 &= 0 \\ 3u_1(u_1 - 1)^2 v_1^3 + 3v_1 - u_2^2 v_2 - 2u_2^3 v_2^3 &= 0 \\ (u_1^2 - 5u_1 + 5)v_1^3 - 3(u_1^3 + 6u_1^2 - 9u_1 + 1)v_1^2 - u_2 v_2^3 - u_2^2 v_2 &= 0 \end{aligned}$$

The points obtained by the trace have the coordinates (u_1, v_1, u_2, v_2) and trace simultaneously the image of the intersection curve in both parameter spaces. The curve in 3-space is recovered from (u_1, v_1) via the coordinate functions $G_{1,j}$, or from (u_2, v_2) via $G_{2,j}$, where $j = 1, 2, 3$. $\diamond$

6.3.3 Surface Operations in Higher Dimensions

Some surface-intersection problems can be expressed straightforwardly as the simultaneous intersection of $n-1$ hypersurfaces in n-dimensional space, where $n > 3$. Given such a formulation, we can use the generalized tracing method. As an illustration, we consider offset surface intersection. Other surface operations may be similarly expressed and treated; see the notes at the end of the chapter.

A number of geometric operations on solid models require *offsetting* a given surface by some distance r. That is, given a surface f, we wish to determine a surface g such that, for every point p of f, there is a point q on g such that the distance between p and q is exactly r, and the line $\overline{(p,q)}$ is perpendicular to f at p.

There are methods for determining an implicit equation for the r-offset g of f. Here, f could be implicit or parametric. However, offsetting may entail considerable symbolic computation, and it may therefore be advantageous to circumvent determining g explicitly, and to reformulate the problem in a higher-dimensional space.

Offset Surface Construction Using Envelopes

Consider a parametric surface f given by

$$\begin{aligned} x &= f_1(s,t) \\ y &= f_2(s,t) \\ z &= f_3(s,t) \end{aligned}$$

Let $\mathbf{n}(s,t) = (n_x(s,t), n_y(s,t), n_z(s,t))$ be the unit normal to f; that is, $\mathbf{n}$ is a vector of length 1. Then, the points

$$\begin{aligned} x &= f_1(s,t) + r n_x(s,t) \\ y &= f_2(s,t) + r n_y(s,t) \\ z &= f_3(s,t) + r n_z(s,t) \end{aligned} \tag{6.6}$$

are on the r-offset of f. The formula (6.6) can be used as the definition of the r-offset of f, but the disadvantage is that this formulation is not algebraic, since $\mathbf{n}$ involves a square root. In fact, examples can be constructed such that the surface described by (6.6) is not algebraic. It is, however, part of an algebraic surface. To find this algebraic surface, we must consider the points at distance r on *both sides* of f. We describe a method for determining the two-sided offset surface.

We consider a family of spheres S of radius r, each of whose centers is constrained to lie on the surface f. The *envelope* of this family contains the set

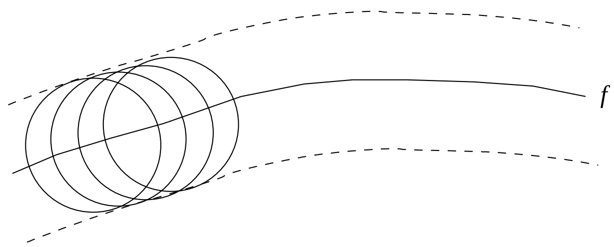


Figure 6.5 Offset Curve as Envelope of Circles

of points whose distance from f is r. Figure 6.5 illustrates this concept in two dimensions. Intuitively, the envelope points are determined by intersecting a sphere in generic position with two adjacent spheres, differentially moved in independent directions on the surface. Using techniques from differential geometry, we can prove that we find these points by solving a system of algebraic equations.

Theorem
The envelope points of a family of surfaces $S(x, y, z, \alpha_1, \alpha_2)$, parameterized by α_1 and α_2, satisfy the three equations

$$S = 0 \quad (6.7)$$
$$\frac{\partial S}{\partial \alpha_1} = 0 \quad (6.8)$$
$$\frac{\partial S}{\partial \alpha_2} = 0 \quad (6.9)$$

The theorem generalizes to all dimensions. In our situation, we apply the theorem as follows. Given the parametric surface f as

$$\begin{aligned} x &= f_1(s,t) \\ y &= f_2(s,t) \\ z &= f_3(s,t)) \end{aligned}$$

we consider the spheres

$$S: \ (x - f_1(s,t))^2 + (y - f_2(s,t))^2 + (z - f_3(s,t))^2 - r^2 = 0$$

Note that $\alpha_1 = s$ and $\alpha_2 = t$. We form the partial derivatives of S by s and by t and obtain

$$S_s: \ -2(x-f_1)\frac{\partial f_1}{\partial s} - 2(y-f_2)\frac{\partial f_2}{\partial s} - 2(z-f_3)\frac{\partial f_3}{\partial s}$$

$$S_t: \ -2(x-f_1)\frac{\partial f_1}{\partial t} - 2(y-f_2)\frac{\partial f_2}{\partial t} - 2(z-f_3)\frac{\partial f_3}{\partial t}$$

By eliminating s and t from the three equations, an algebraic description of the offset surfaces is obtained.

For implicit f, equations (6.8) and (6.9) should be replaced with the directional derivatives in two independent tangent directions on the surface f, and an additional equation is needed that expresses that the centers of the sphere must lie on the implicit surface. The pattern is as follows:

$$\begin{aligned} S: \ (x-u_1)^2 + (y-u_2)^2 + (z-u_3)^2 - r^2 &= 0 \\ f(u_1, u_2, u_3) &= 0 \\ \nabla_u S \cdot \mathbf{t}_1 &= 0 \\ \nabla_u S \cdot \mathbf{t}_2 &= 0 \end{aligned} \tag{6.10}$$

Here, $\mathbf{t}_1$ and $\mathbf{t}_2$ are two linearly independent tangent vectors to f at the point (u_1, u_2, u_3), and

$$\nabla_u S = \left(\frac{\partial S}{\partial u_1}, \frac{\partial S}{\partial u_2}, \frac{\partial S}{\partial u_3}\right)$$

Elimination of u_1, u_2, and u_3 from the set of equations (6.10) results in an implicit equation describing the offset of f.

We could now eliminate s and t in the parametric case, or eliminate u_1, u_2, and u_3, in the implicit case. The result, in each case, is an implicit equation for the offset surface which then is intersected with some other surface, say g. However, the symbolic computations incurred by the elimination step could be forbidding. So, we will intersect g with the system of equations describing the offset, thus tracing the intersection in a dimension higher than three. Example 6.2 illustrates the method. Such a trace derives the following additional information:

- In the parametric case, each point is traced in five dimensions, and has the coordinates (x, y, z, s, t). Here, $p = (x, y, z)$ is the point on the intersection of the offset surface of f with g. The point (s, t) in parameter space determines the *footpoint* of p; that is, the point on f at distance r from p.

- In the implicit case, each point is traced in six dimensions, and has the coordinates (x, y, z, u_1, u_2, u_3). Again, $p = (x, y, z)$ is the point on the offset surface intersection with g, and (u_1, u_2, u_3) is its footpoint on f.

Note that we can intersect two offset surfaces with each other by combining the respective systems of equations. This raises the dimensionality of the problem, but not its difficulty.

It is important to note the following points about the envelope method for formulating offsets:

1. The offset surface may self-intersect. In applications, self-intersections are undesirable, and "interior" surface parts may be obtained that one wants to eliminate. See Figure 6.6 for an illustration in two dimensions. Neither formulation (6.6) nor the envelope method will automatically eliminate those interior parts.

2. In constructing an algebraic description of offset surfaces, we operate implicitly over the field of complex numbers, and we obtain certain surface components at infinity. Both result in additional points that are described by the equations.[5] Figure 6.7 shows an example in two dimensions. Those points are not generated in formulation (6.6). On the other hand, that formulation cannot handle singularities.

In our view, these phenomena are due to the fact that algebraic computations implicitly require projective spaces over an algebraically closed ground field, as explained in Section 7.2.1 in Chapter 7. Insisting on working in real affine spaces substantially reduces the available mathematical machinery.

Example 6.2: Consider the ellipsoid $f = 2x^2 + 9y^2 + 18z^2 - 18$. We plan to intersect its offset by 1 with a cylinder $h = (x-3)^2 + y^2 - 1$. We consider a sphere of radius 1 centered at the point $p = (u_1, u_2, u_3)$ of the ellipsoid:

$$\begin{aligned} S:\ (x-u_1)^2 + (y-u_2)^2 + (z-u_3)^2 - 1 &= 0 \\ 2u_1^2 + 9u_2^2 + 18u_3^2 - 18 &= 0 \end{aligned}$$

Here, the second equation ensures that p lies on the ellipsoid. To find the directional derivatives of S, we must determine two independent tangent directions to the ellipsoid at p. Now the gradient at p is

$$\nabla f = (4u_1, 18u_2, 36u_3) = 2(2u_1, 9u_2, 18u_3)$$

[5]In the case of parametric curves, some of these additional points can be eliminated by dividing out certain factors. See also the notes at the end of the chapter.

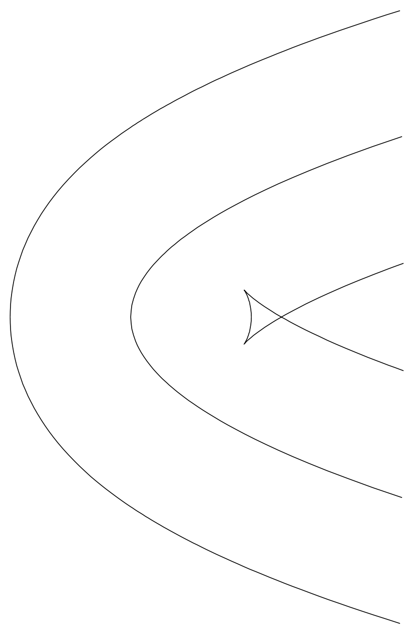


Figure 6.6 Interior Offset Part at Self-Intersections

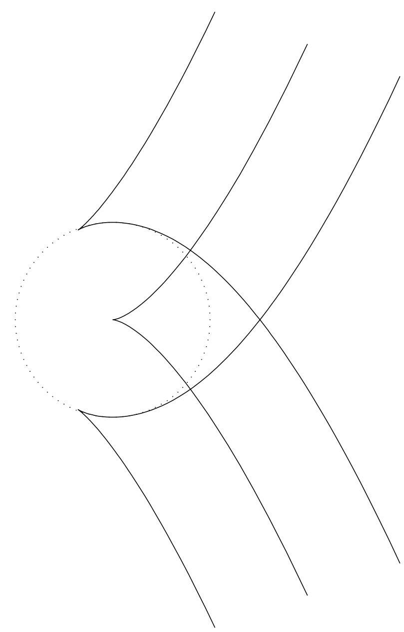


Figure 6.7 Additional Offset Points at Singularity

Hence, perpendiculars to ∇f will be tangent directions. We choose

$$\begin{aligned} \mathbf{t}_1 : &\quad (-9u_2, 2u_1, 0) \\ \mathbf{t}_2 : &\quad (0, -18u_3, 9u_2) \end{aligned}$$

and observe that $\mathbf{t}_1 \perp \nabla f$, $\mathbf{t}_2 \perp \nabla f$, and that $\mathbf{t}_1$ and $\mathbf{t}_2$ are linearly independent when $u_2 \neq 0$. Moreover,

$$\nabla_u S = 2(-(x-u_1), -(y-u_2), -(z-u_3))$$

Equations (6.8) and (6.9) specialize to

$$\begin{aligned} \nabla_u S \cdot \mathbf{t}_1 &= 9(x-u_1)u_2 - 2(y-u_2)u_1 \\ \nabla_u S \cdot \mathbf{t}_2 &= 18(y-u_2)u_3 - 9(z-u_3)u_2 \end{aligned}$$

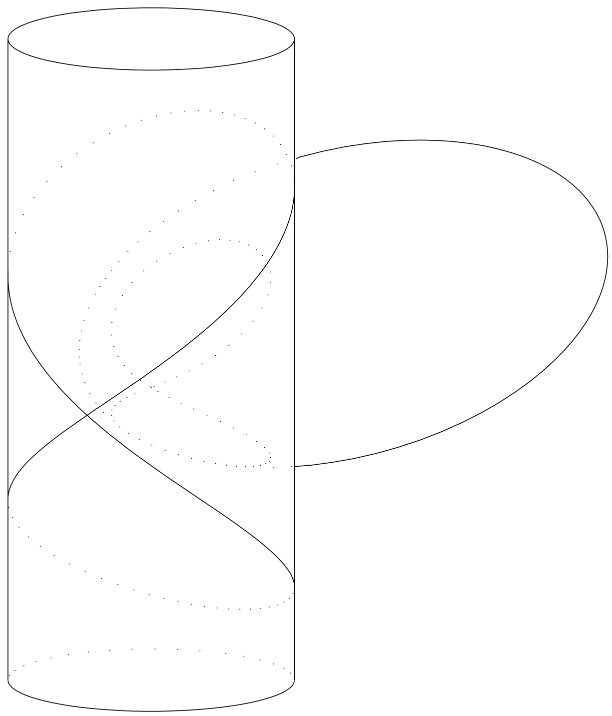


Figure 6.8 Intersecting a Cylinder with an Offset of an Ellipsoid

Therefore, the intersection with h is described by the equations

$$\begin{aligned}(x-u_1)^2+(y-u_2)^2+(z-u_3)^2-1 &= 0\\ 2u_1^2+9u_2^2+18u_3^2-18 &= 0\\ 9(x-u_1)u_2-2(y-u_2)u_1 &= 0\\ 18(y-u_2)u_3-9(z-u_3)u_2 &= 0\\ (x-3)^2+y^2-1 &= 0\end{aligned} \tag{6.11}$$

These are five equations in the six unknowns x, y, z, u_1, u_2, u_3. Here, x, y, and z are the coordinates, in 3-space, of the intersection curve of the offset of f with g. Moreover, $q = (u_1, u_2, u_3)$ is the footpoint of $p = (x, y, z)$ on f; that is, it is a point on f such that the surface normal through q passes through p, and such that the Euclidian distance $\overline{(p,q)}$ is the offset distance. Rather than eliminating the unknowns u_1, u_2, and u_3, we trace the intersection curve in six dimensions, tracking simultaneously the intersection of the offset with g, as well as the footpoint curve on f. Figure 6.8 shows this trace. $\diamond$

6.4 Mapping Surface Intersections to Plane Curves

A general approach to surface-intersection evaluation is to map the surface intersection to a plane curve $h(u, v) = 0$. The approach is appealing for a number of reasons. For one, a plane curve can be traced through singularities, as explained later in this chapter. An analogous process for surface intersection could be devised in principle, but it would be substantially more complex because it would have to map the intersecting surfaces simultaneously such that the singularity of their intersection would be resolved.

On the other hand, the mapping approach has to face a number of difficulties that reduce its attractiveness. These include the cost of constructing the map, the numerical inaccuracies that might arise in the substitution process, and, finally, the high degree of h, which is, in general, the product of the surface degrees and usually leads to numerical difficulties. We discuss several techniques that can be applied in various situations. None of them avoid all the problems mentioned.

6.4.1 Substitution Maps

In a substitution map, we substitute the parametric form of one surface into the implicit form of the other, thereby obtaining a plane curve in the parameter space of the first surface. If we intersect a parametric with an implicit surface, the cost of constructing the map is just the cost of doing the substitution. Otherwise, we must add the cost of converting the representation of one of the surfaces from parametric to implicit, or vice versa. In most of those situations, the cost of the representation conversion will dominate.

Intersecting Two Parametric Surfaces

When intersecting two parametric surfaces, we implicitize one of them. Implicitization is always possible, and can be done either by resultant computations or by Gröbner bases techniques. The resultant-based computation suffers from the extraneous factor problem. If the implicitized surface is

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

and the parametric surface is

$$\begin{aligned} x &= g_1(u, v) \\ y &= g_2(u, v) \\ z &= g_3(u, v) \end{aligned}$$

then the plane algebraic curve is

$$h(u,v) = f(g_1(u,v), g_2(u,v), g_3(u,v)) = 0$$

We can then trace $h = 0$ in u, v space, and map each point via the rational functions g_i.

Example 6.3: Consider the intersection of the parametric surfaces

$$f : \begin{cases} x = st \\ y = st^2 \\ z = s^2 \end{cases} \quad \text{and} \quad g : \begin{cases} x = u^2 - v^2 \\ y = 2uv \\ z = u^2 + v^2 \end{cases}$$

We implicitize the surface f using the Gröbner bases method discussed in Section 7.5.1 of Chapter 7. and obtain the implicit form

$$x^4 - y^2 z = 0$$

Substitution of the second surface into this implicit form yields the plane curve

$$h : \ (u^2 - v^2)^4 - 4u^2v^2(u^2 + v^2) = 0$$

Each point (u, v) of h corresponds to the point $(u^2 - v^2, 2uv, u^2 + v^2)$ on g, which must also be on f; hence, it is a point on the intersection. A trace of h is shown in Figure 6.9. $\diamond$

Intersecting Two Implicit Surfaces

When intersecting two implicit surfaces, we would like to parameterize one of them. Not every implicit surface possesses a rational parametric form, however, so this approach needs to be modified. It can be shown that the intersection of two implicit surfaces always lies on a parameterizable surface. That is, given the surfaces

$$\begin{aligned} f(x,y,z) &= 0 \\ g(x,y,z) &= 0 \end{aligned}$$

there is a surface h that is parameterizable and contains the intersection of f and g. The surface has the form

$$h(x,y,z) = h_1 f + h_2 g = 0$$

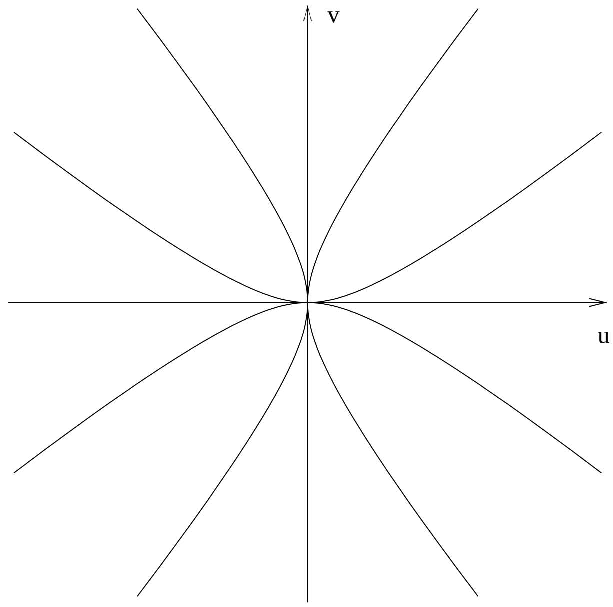


Figure 6.9 Trace of $(u^2 - v^2)^4 - 4u^2v^2(u^2 + v^2) = 0$

where the coefficients h_1 and h_2 are polynomials. The computation for obtaining h is conceptually simple, as is the parameterization, since h will be a monoid whose singular point we will know.

We describe the derivation of h. First, we *homogenize* f and g, obtaining $F(w, x, y, z)$ and $G(w, x, y, z)$. As long as $w \neq 0$, the curve $F \cap G$ is identical to $f \cap g$. We select one of the variables as *main variable*, and rewrite F and G as polynomials in this variable, say w:

$$\begin{aligned} F &= u_n w^n + u_{n-1} w^{n-1} + \cdots + u_1 w + u_0 \\ G &= v_{n'} w^{n'} + v_{n'-1} w^{n'-1} + \cdots + v_1 w + v_0 \end{aligned}$$

Without loss of generality, we can assume that $n \geq n' > 1$, and determine the polynomials

$$\begin{aligned} F_1 &= u_n w^{n-n'} G - v_{n'} F \\ G_1 &= (u_0 G - v_0 F)/w \end{aligned}$$

In effect, in F_1, we cancel the highest terms in w; in G_1, we cancel the lowest term. Note that both F_1 and G_1 contain the intersection curve of F and G, since they are algebraic combinations of the two surfaces.

Both F_1 and G_1 have degree at most $n - 1$ in w. If one of them is linear in w, then we stop; we have found the desired surface. If neither is linear, then we repeat the calculation using F_1 and G_1 in place of F and G. Since at each step the maximum degree in w is lowered by at least one, the computation derives the desired monoid equation after at most n steps, in the form

$$wH_{m-1}(x, y, z) + H_m(x, y, z) = 0$$

This surface is then parameterized by

$$\begin{aligned} w(u, v, s) &= -H_m(u, v, s)/H_{m-1}(u, v, s) \\ x(u, v, s) &= u \\ y(u, v, s) &= v \\ z(u, v, s) &= s \end{aligned}$$

as described in Section 5.5.4 of Chapter 5. The parametric forms are now substituted into the equation of G and give a plane curve in homogeneous form. After dehomogenizing, this is the desired plane curve.

Example 6.4: Consider the intersection curve of the cylinder $f = x^2 + (z+1)^2 - 1 = 0$ and the sphere $g = x^2 + y^2 + (z+2)^2 - 4 = 0$. Homogenizing, we obtain $F = x^2 + z^2 + 2zw$ and $G = x^2 + y^2 + z^2 + 4zw$. The intersection curve is an irreducible degree-4 space curve with a nodal singularity at the origin. We select z as the main variable. Accordingly, we compute

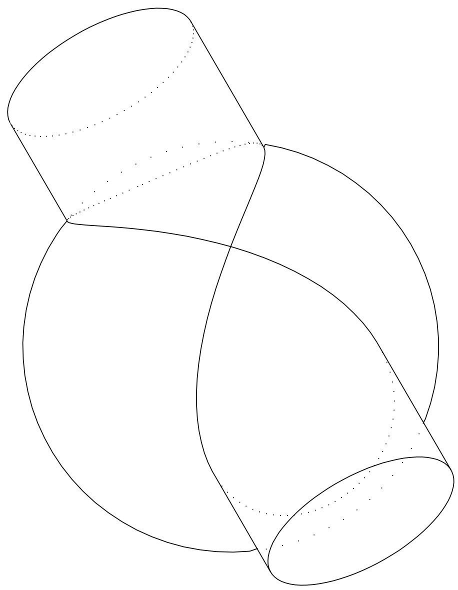
$$\begin{aligned} F_1 &= G - F = y^2 + 2zw \\ G_1 &= [(x^2 + y^2)F - x^2 G]/z = y^2 z + 2(y^2 - x^2)w \end{aligned}$$

Both polynomials are linear in z. F_1 is simpler and has the parameterization

$$\begin{aligned} z &= -\frac{s^2}{2u} \\ w &= u \\ x &= v \\ y &= s \end{aligned}$$

Substitution into G yields the plane curve

$$s^4 + 4u^2(v^2 - s^2) = 0$$



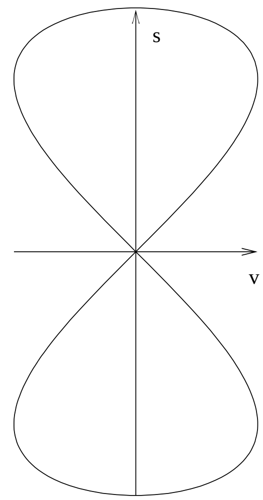


Figure 6.10 Space Curve and Its Planar Image in Parameter Space of Monoid

Dehomogenizing with $u = 1$ yields $s^4 - 4(v^2 - s^2) = 0$. Both the space curve and its planar image are shown in Figure 6.10. $\diamond$

Example 6.4 is favorable because the degree of the plane curve obtained is the minimum degree possible. In general, the monoid method will introduce extraneous factors and will yield plane curves of higher degree than needed. An example of this phenomenon is easily constructed.

We intersect the torus

$$(x^2 + y^2 + z^2 - w^2)^2 + 8w^2(z^2 - x^2 - y^2 - w^2) + 16w^4 = 0$$

with the ellipsoid

$$36(x - w)^2 + 4(y - w)^2 + 9z^2 - 36w^2 = 0$$

We apply the monoid construction,[6] and obtain in three steps a rational surface of degree 8. The respective degrees obtained in each step are as follows:

$$\begin{array}{lccc} F_i \text{ degrees}: & 4 & 6 & 8 \\ G_i \text{ degrees}: & 5 & 4 & 9 \end{array}$$

[6]In practice, one should parameterize the ellipsoid or the torus, since both are rationally parameterizable.

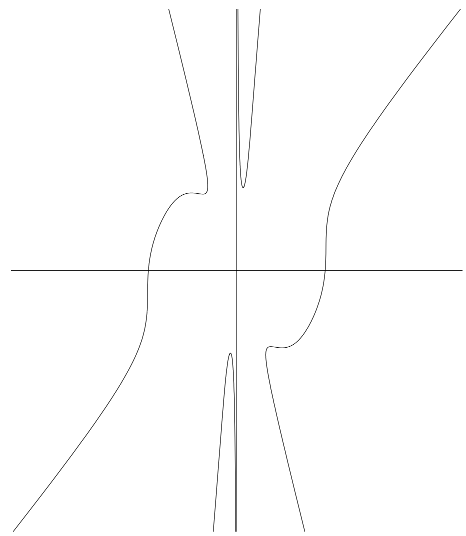


Figure 6.11 Ellipsoid/Torus Intersection in Parameter Space

Note that the degree drop in the second stage is due to a common factor of degree 4 of the w^0 coefficients.

We use the monoid of degree 8 and substitute its parametric form into the equation of the ellipsoid. This yields a plane curve of degree 16 that factors. The curve has three components, of degree 2, 6, and 8, respectively. The degree-2 component is $36x^2 + 4y^2 + 9z^2$, and is the intersection of the ellipsoid with the plane at infinity. We can verify that this curve is not on the torus, and conclude that the degree-2 component is an extraneous factor. By Bezout's theorem, we expect an intersection curve of degree 8. In conjunction with the rejection of the degree-2 component, this implies that the degree-6 component is also extraneous, so the degree-8 component is the sought curve. It is shown in Figure 6.11.

6.4.2 Projection Methods

The second general approach to mapping a space curve to a plane curve is to use projection. In principle, the construction of these maps is straightforward. The main problem, however, is that the point from which to project must be chosen carefully: A poorly chosen point will result in a map that cannot be inverted. Such a projection map would not permit mapping the points of the plane curve back to space-curve points, so that the plane-curve trace would yield no information.

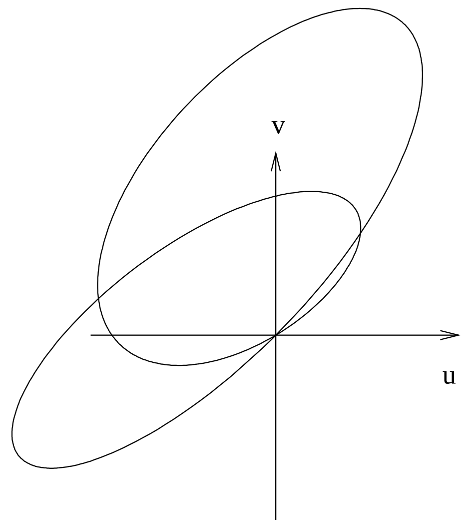


Figure 6.12 Projection of the Cylinder/Sphere Intersection

If we assume that the two surfaces intersect transversally — that is, that the surface gradients are linearly independent almost everywhere — and that the curve itself is irreducible, then it can be shown that a good projection point can be chosen by the following computation:

1. Transform the surface equations by a nonsingular linear transformation with symbolic coefficients.

2. Project the intersection by a resultant computation.

3. Choose random numeric values for the coefficients and verify that the projection does not degenerate.

Note that step 3 succeeds with a probability of one. Almost all assignments will result in a nonsingular linear transformation. Moreover, assignments failing to produce a good projection are in directions at which infinitely many curve-point pairs line up. Those directions constitute a ruled surface, and any view point not on that surface yields a suitable projection.

Example 6.5: We consider the surfaces of Example 6.4, intersecting

$$\begin{aligned} f &: \quad x^2 + (z+1)^2 - 1 \\ g &: \quad x^2 + y^2 + (z+2)^2 - 4 \end{aligned}$$

Instead of substituting with symbolic coefficients a_{ij}, we substitute

$$\begin{aligned} x &= u+v+w \\ y &= u+2v+4w \\ z &= u-2v-w \end{aligned}$$

For random values, we expect that none of the a_{ij} are zero, and that none of the coefficient expressions vanish, after substitution. This is also the case for this substitution, although it is not based on random values.

With this substitution, we obtain the polynomial $F(u,v,w)$ from f, and the polynomial $G(u,v,w)$ from g. We eliminate one of the variables — say, w — obtaining a polynomial $H(u,v) = Res_w(F,G)$, where

$$\begin{aligned} H(u,v)/4 &= 400v^4 - 960uv^3 - 952v^3 + 1256u^2v^2 + 1848uv^2 + 490v^2 \\ &\quad -816u^3v - 1554u^2v - 784uv + 289u^4 + 588u^3 + 294u^2 \end{aligned}$$

Recall that the initial form of a polynomial consists of the terms of lowest degree. Consequently, the initial form of the curve H is $h_0 = 490v^2 - 784uv + 294u^2$. The initial form factors $h_0 = 98(v-u)(5v-3u)$; that is, the singularity of H at the origin is a node.

The curve H is shown in Figure 6.12. It has three singularities. The one at the origin corresponds to the singularity of the surface intersection. The other two have been introduced by the projection.

We stated that the projection will succeed with high probability. As example of a poor projection point, consider the assignment

$$x = w, \qquad y = v, \qquad z = u$$

Here we obtain the plane curve

$$(v^2 + 2u)^2$$

That is, we obtain a double parabola. There is no rational map from this plane curve to the space curve, so tracing the parabola would be useless. $\diamond$

6.5 Plane Algebraic Curves

We now consider how to trace a plane algebraic curve $f(x,y) = 0$. Tracing a plane algebraic curve is fundamental, because every algebraic space curve can be mapped birationally to a plane algebraic curve. This observation has been used in various ways in the surface-intersection problem, and continues to be researched as an approach to intersection evaluation. The method to be described can trace through curve singularities of arbitrary structure.

As before, the bulk of the tracing will be done numerically, and the routines continue to be structured as before:

1. Construct a local approximant at the curve point p.

2. With a selected step size, derive a new curve-point estimate q.

3. Refine the estimate q iteratively, obtaining a curve point.

The numerical tracing routine performs very well, except at singularities. All purely numerical tracing routines fail at a singular curve point for the same reason: Technically, the routines depend on the underlying assumption that there exists a system of linear equations that determines the local structure of the curve with sufficient accuracy. In our case, this was the system of equations (6.1). At a singular point, however, these equations are nonlinear, as explained later, and approaching the problem as a linear one would be inappropriate. Hence, we seek methods for analyzing singularities.

Rather than dealing directly with nonlinear equations, we will apply a classical result from algebraic geometry that states that every algebraic curve $f(x, y) = 0$ can be transformed birationally into a curve $g(x, y) = 0$ that is devoid of singularities. Thus, we plan to trace g in the vicinity of singular points of f, and to map the points of g back to corresponding points of f.

It would be nice if we needed to trace only g. Unfortunately, g cannot be so used, since we might have to pass through infinity. So, we trace f whenever possible and trace only the critical segments of f on g — the segments containing singularities.

6.5.1 Place of a Curve

In this section, we define the notion of a *place* of an algebraic curve. We need this concept to analyze the nature of singularities and to elucidate the effect of quadratic transformations used by the tracing algorithm to resolve singularities.

Definition of Place of a Curve

At the point $p = (a_0, b_0)$ of the plane algebraic curve $f(x, y) = 0$, we define the formal power series

$$\begin{aligned} x(s) &= a_0 + a_1 s + a_2 s^2 + \cdots \\ y(s) &= b_0 + b_1 s + b_2 s^2 + \cdots \end{aligned} \tag{6.12}$$

and require that $f(x(s), y(s)) \equiv 0$. We think of the pair as a *local parameterization* of f. It is called a *place* of f at p, and exists because of Newton's theorem.

Newton's theorem generalizes the implicit function theorem: The implicit function theorem states that, for a regular curve point p at which the partial derivative f_y is not zero, there exists a neighborhood U in which an analytic function $y = h(x)$ can be defined such that $f(x, h(x)) \equiv 0$, for all x in U. By introducing a new variable s and defining two analytic functions $x = h_1(s)$ and $y = h_2(s)$, such that $f(h_1(s), h_2(s)) \equiv 0$, the hypothesis $f_y \neq 0$ can be abolished. Furthermore, by allowing possibly more than one pair of functions of the form of equation (6.12) at p, the assumption that p is not singular can be removed. So generalized, we obtain Newton's theorem, which says, roughly, that given a polynomial $f(x, y)$ and a point $p = (a_0, b_0)$ on it, there exist power series of the form of equation (6.12) such that

$$f(x(s), y(s)) \equiv 0 \tag{6.13}$$

These power series can be viewed as formal series, in which case equation (6.13) is an algebraic identity; or they can be considered as defining analytic functions, in which case their convergence properties must be considered as well. In the following discussion, we adopt the former point of view.

The notion of place is more specific than that of a curve point. At a regular curve point p, the curve f has only one place, and that place can be shown to be essentially the Taylor expansion of f at p. At singular points, the curve may have several places. When considered within the disk of convergence, a place is simply an analytic curve branch.

Basic Properties of Places and Singularities

A place is *regular* if a_1 and b_1 are not simultaneously zero; otherwise, it is *singular*. We say that the place is *centered* at (a_0, b_0). For example, the place

$$\begin{aligned} x(s) &= s \\ y(s) &= s + \frac{1}{2}s^2 - \frac{1}{8}s^3 \pm \ldots \end{aligned}$$

is regular, whereas the place

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

is singular. If a curve has exactly one regular place centered at (a_0, b_0), then (a_0, b_0) is a regular curve point. Otherwise, it is a singular curve point. Thus, a singular curve point is one at which the curve has either one singular place or at which the curve has two or more places. In the latter case, none, some, or all of the places could be singular as well.

The definition of a singular curve point in terms of places can be shown to be equivalent to the definition in terms of vanishing partial derivatives. We define the *derivative* $h'(s)$ of the series $h(s) = a_0 + a_1 s + a_2 s^2 + a_3 s^3 + \cdots$ as

$$h'(s) = a_1 + 2a_2 s + 3a_3 s^2 + \cdots$$

Then, it can be shown that at the regular curve point (a_0, b_0) the partial derivatives of f are proportional to the derivatives of the place; that is,

$$\begin{aligned} f_x(a_0, b_0) &= \alpha x'(0) \\ f_y(a_0, b_0) &= \alpha y'(0) \end{aligned}$$

where $\alpha \neq 0$.

The *order* of a power series $a_1 s + a_2 s^2 + a_3 s^3 + \cdots$ is the minimum index k such that $a_k \neq 0$. Similarly, the *order* of a place centered at the origin is the smallest index k such that a_k and b_k are not both zero. The order of a regular place is always one, the order of a singular place is always greater than one. A place

$$\begin{aligned} x(s) &= a_1 s + a_2 s^2 + \cdots \\ y(s) &= b_1 s + b_2 s^2 + \cdots \end{aligned}$$

intersects a curve $g(x, y) = 0$ at the origin with *multiplicity* k if k is the order of the power series $g(x(s), y(s))$. Intersection multiplicity is also called *order of contact*.

Example 6.6: The place $(x(s) = s^2, y(s) = s^3)$ intersects the line $x + y$ with multiplicity 2, since $x(s) + y(s) = s^2 + s^3$ has order 2. However, the line $y = 0$ intersects the place with multiplicity 3. $\diamond$

Let $P_1 = (x_1(s), y_1(s))$ and $P_2 = (x_2(s), y_2(s))$ be two places centered at the origin. We would like to define their intersection multiplicity at the origin as k whenever the first $k + 1$ coefficients of $x_1(s)$ and $x_2(s)$ and of $y_1(s)$ and $y_2(s)$ agree. This problem is not quite so simple, because there are different ways to write the power series. Moreover, since the line $x + y = 0$ has the place $(x(s) = s, y(s) = s)$ at the origin, such a definition would not be compatible with the intersection multiplicity of places with curves. The proper definition requires reparameterizing the two places such that a canonical form is obtained, after which the multiplicity — or, equivalently, the order of contact — can be defined as the order of a certain power series. The details are omitted.

A key theorem from algebraic geometry states that two places either are in contact of finite order, or else are equal. Later on, we will use this theorem to show that the different places centered at the same singular point of a curve can be "separated." Such a separation is one of the key aspects of resolving curve singularities. Note that the theorem is reminiscent of Bezout's theorem.

If we think of a place as an analytic branch, then it makes sense to define its tangent. Intuitively, a tangent to the place

$$\begin{aligned} x(s) &= a_0 + a_1 s + a_2 s^2 + \cdots \\ y(s) &= b_0 + b_1 s + b_2 s^2 + \cdots \end{aligned}$$

is a line through (a_0, b_0) that intersects the place with a higher multiplicity; that is, it is in higher order of contact with the place than almost all other lines.

Example 6.7: Consider the place $(x(s) = s^2, y(s) = s^3)$, centered at the origin. Let

$$ux + vy = 0$$

be a line through the origin, where u and v are not both zero. We have

$$ux(s) + vy(s) = us^2 + vs^3$$

So, the line intersects the place with multiplicity 2, except when $u = 0$, for then the intersection multiplicity is 3. Therefore, $y = 0$ is tangent to the place (s^2, s^3). $\diamond$

It can be proved that there is exactly one tangent to a place. Moreover, it can be proved that at a regular curve point the tangent to the place is the curve tangent, and that at a singular point p the curve tangents consist of the tangents to the places of the curve that are centered at p. In particular, if the origin is a point on f, then the tangent lines to f at the origin are the linear factors of the initial form of f. Thus, the notion of tangency to a place is more specific than is the concept of tangent space introduced in Section 5.3.1 of Chapter 5.

A Method for Computing Places

Given a point $p = (a_0, b_0)$, one way to determine the place(s) of $f(x, y)$ is to set up the series of equations (6.12) formally, to substitute them into f, and to set the coefficients of the resulting power series to zero. For example, let $f = y^2 - x^2 - x^3$, and consider the point $(0, 0)$. We substitute $\sum_{i \geq 1} a_i s^i$ for x, and $\sum_{i \geq 1} b_i s^i$ for y, and set the coefficient of each power of s to zero. This yields the following system

$$
\begin{aligned}
a_1^2 - b_1^2 &= 0 \\
2b_1b_2 - 2a_1a_2 - a_1^3 &= 0 \\
2b_1b_3 + b_2^2 - 2a_1a_3 - a_2^2 - 3a_1^2a_2 &= 0 \\
2b_1b_4 + 2b_2b_3 - 2a_1a_4 - 2a_2a_3 - 3a_1^2a_3 - 3a_1a_2^2 &= 0 \\
&\vdots
\end{aligned}
\tag{6.14}
$$

which has the solutions

$$
\begin{aligned}
x(s) &= s \\
y(s) &= s + \frac{1}{2}s^2 - \frac{1}{8}s^3 \pm \cdots
\end{aligned}
$$

and

$$
\begin{aligned}
x(s) &= s \\
y(s) &= -s - \frac{1}{2}s^2 + \frac{1}{8}s^3 \mp \cdots
\end{aligned}
$$

Each solution is a distinct regular place of f at the origin. Since f has more than one place, it is singular at $(0, 0)$, which is also evident from the nonlinear initial form $y^2 - x^2 = (y - x)(y + x)$ and from the graph of the curve shown in Section 6.5.3 in Figure 6.14 on the left.

The system of equations derived in this way agrees formally with the system of equations (6.1), formulated in Section 6.2. The connection becomes evident when considering the derivatives of a place

$$
\begin{aligned}
x'(s) &= a_1 + 2a_2s + 3a_3s^2 + \dots \\
y'(s) &= b_1 + 2b_2s + 3b_3s^2 + \dots
\end{aligned}
$$

Higher-order derivatives are defined analogously. Then, the system of equations (6.14) can be shown to be of the form

$$
\nabla f \cdot \mathbf{r}^{(m)} = b_{f,m}
$$

with $\mathbf{r} = (x(s), y(s))$. Thus, the difference between the regular and the singular case is simply that in the regular case this system is linear, whereas in the singular case it is nonlinear.

Places on Space Curves

The notion of place generalizes directly to higher-dimensional algebraic curves. As before, the system of equations (6.1) can be formulated and solved for each space-curve point. For regular curve points, the system is linear; for singular curve points, it is nonlinear. When solving such a system, the following theorem is helpful.

Theorem
Let $(x_1(s), x_2(s), ..., x_n(s))$ be a place of an algebraic space curve. Then the parameter s can be chosen such that one of the coordinates x_j has the form

$$x_j(s) = s^k$$

For a regular curve point, k will be 1, so this theorem specializes to the implicit function theorem.

6.5.2 Quadratic Transformations

Let $f(x, y) = 0$ be a plane algebraic curve on which the origin $(0, 0)$ is a singular point. We wish to construct a birational transformation τ from the curve f to a curve g with the following properties. The transformation is bijective in a neighborhood of the origin, except, possibly, at the origin.[7] Moreover, each place of f centered at the origin is mapped to a regular place of g, and the center of each such place is a regular point of g.

Intuitively, τ separates all places of f at the origin, and, if any one of them is a singular place, it is transformed by τ into a nonsingular place centered at a regular curve point of g. Thus, τ "resolves" the singularity into several regular curve branches situated at different nonsingular points. Note that we consider only singularities at the origin. This is sufficient, because a singular curve point can always be brought to the origin by a change of coordinates.

The birational map effecting a resolution of the singularity is constructed incrementally from two quadratic transformations T_1 and T_2:

$$\begin{aligned} T_1: \quad & x_1 = x \\ & y_1 = y/x \end{aligned}$$

$$\begin{aligned} T_2: \quad & x_2 = x/y \\ & y_2 = y \end{aligned}$$

[7] Strictly speaking, τ is not bijective on certain lines. When constructing τ, we will ensure that these "exceptional lines" are not tangent to the branch we trace.

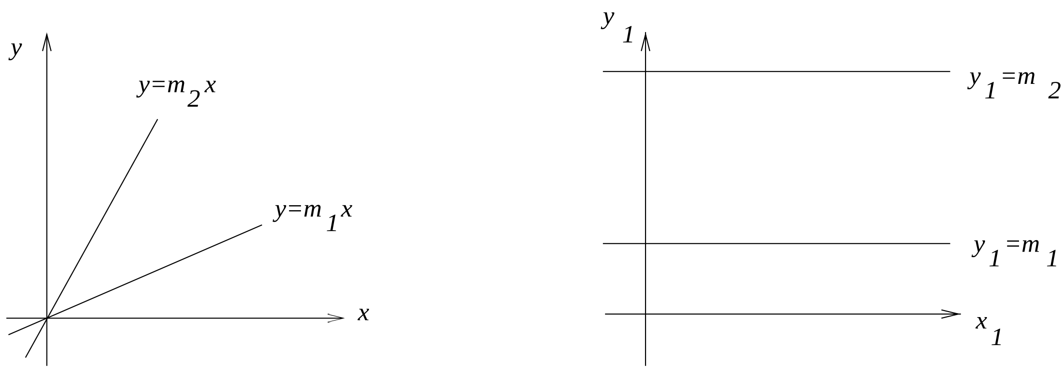


Figure 6.13 Quadratic Transformation $T_1 : (x, y) \to (x, y/x)$

Initially, we restrict attention to T_1. Its inverse is evidently $x = x_1$ and $y = x_1 y_1$. The basic properties of T_1 are as follows:

1. T_1 maps the set $\{(x, y) \mid x \neq 0\}$ bijectively onto the set $\{(x_1, y_1) \mid x_1 \neq 0\}$.
2. The points $(0, y)$ with $y \neq 0$ are mapped to infinity in the (x_1, y_1)-plane.
3. As we approach the origin on a curve branch, the limit of the image points is the image of the origin on the branch. This limit depends on the direction of approach: If the branch has a tangent with slope m at the origin, then that branch will intersect the y_1 axis in $(0, m)$.

Figure 6.13 shows the effect of T_1 on select lines. The coordinate lines $x = m$ are mapped to $x_1 = m$ for $m \neq 0$. The coordinate lines $y = m$ are mapped to the hyperbolas $y_1 x_1 = m$. Finally, the lines $y = mx$ are mapped to the lines $y_1 = m$ for $m \neq 0$. The effect of T_2 is analogous.

6.5.3 Branch and Curve Desingularization

Assume that $f(x, y) = 0$ has a singularity of order k at the origin, and that the initial form of f is not y^k. When T_1 is applied to f, then the *total transform* of f is $x_1^k g(x_1, y_1) = 0$. The lines $x_1^k = 0$ are not of interest, and we consider $g(x_1, y_1) = 0$ as the *proper transform* of f. In favorable cases, the points of g corresponding to the origin of the (x, y)-plane are not singular. Figures 6.14 and 6.15 show two examples. In more complicated situations, there are singularities at the corresponding points of g, but the structure of these singularities has been simplified in some sense. This statement can

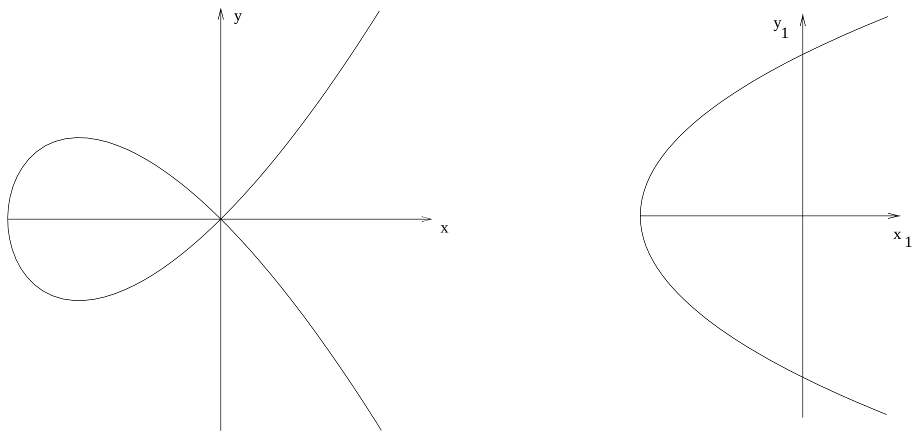


Figure 6.14 Resolution of a Nodal Singularity

be made precise, but requires considerable mathematical machinery, so we restrict our exposition to a somewhat simplistic but intuitive version.

Intuitively, then, there are a number of places at the singularity at the origin that are mapped as follows by T_1: Two places of contact order k will be mapped to two places with order of contact at most $k-1$. Moreover, the image of a singular place will be a place that, if still singular, has a singularity that is structurally simpler. Remarkably, singular places become regular after finitely many applications of quadratic transformations T_1 and/or T_2. Moreover, since different branches cannot have infinite-order contact, they must separate after finitely many quadratic transformations.

So far, we have not commented on the use of T_2. Briefly, if a place has the y axis as tangent, T_1 maps it to infinity where it cannot be further analyzed

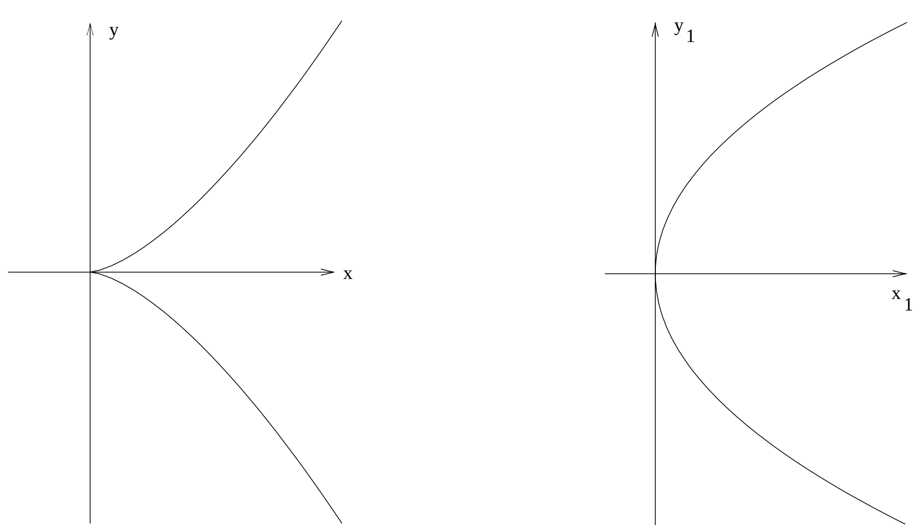


Figure 6.15 Resolution of a Cuspidal Singularity

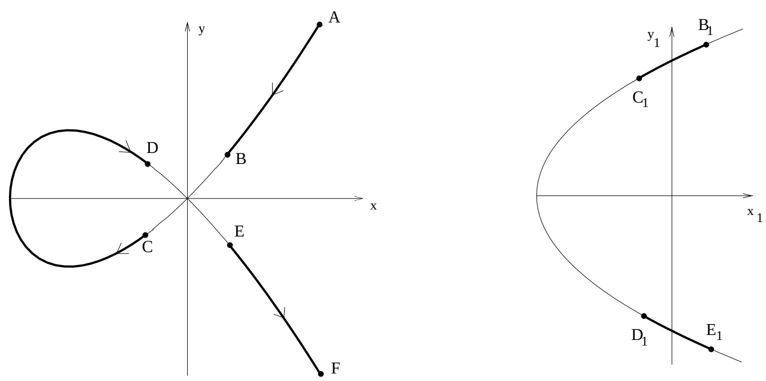


Figure 6.16 Tracing with Desingularization

without passing to the projective plane. Thus, we transform such a branch with the help of T_2.

6.5.4 Tracing with Desingularization

The idea of tracing with desingularization is as follows:

1. Beginning at a regular curve point p, we trace f using the numerical procedure outlined previously.

2. When approaching a singularity q, the numerical trace is suspended at a point r prior to reaching q. Then f is translated such that q becomes the origin.

3. Depending on the tangent direction of the branch we are currently tracing, we transform f to g with T_1 or T_2. Then, beginning at the point r_1 of g corresponding to r, we trace g until we have crossed the singularity, mapping the points on g to f by the inverse of T_1 or T_2.

See also Figure 6.16 for an illustration of the idea. This procedure must be implemented recursively, since a single quadratic transformation may not suffice to resolve the singularity. The major practical concerns are locating the singularity while tracing, and accounting for the intended direction of the trace. Since we do a coordinate transformation to bring the singularity to the origin, we also have to cope with imprecise coordinates of the singularity.

6.5.5 Locating Singularities

When we are numerically tracing f, an impending singularity is detected from the condition number of the matrix

$$\begin{pmatrix} f_x & -f_y \\ f_y & f_x \end{pmatrix}$$

The singular point is the simultaneous intersection of $f = 0$, $f_x = 0$, and $f_y = 0$, and can be found iteratively or by direct methods.

Least-Squares Approach

An iterative approach can be based on a least-squares formulation as follows: Beginning with a nearby curve point p_0, we construct a sequence of points $p_0, p_1, p_2, \ldots$ converging to the singularity. Let $p_{i+1} = p_i + (\delta_x, \delta_y)$. Then we solve the linear system

$$\begin{pmatrix} f_x & f_y \\ f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix} = - \begin{pmatrix} f \\ f_x \\ f_y \end{pmatrix}$$

We rewrite this system in matrix notation as

$$A\Delta = \mathbf{b}$$

where $\Delta = (\delta_x, \delta_y)^T$. This overconstrained system corresponds to the least-squares problem

$$A^T A\Delta = A^T \mathbf{b}$$

For higher-order singularities, higher-order partials may also vanish. Thus, if A does not have full rank, we extend the system by adding, for each vanishing partial h, the equation

$$h_x\delta_x + h_y\delta_y = -h$$

In this manner, a matrix $A^T A$ of full rank is obtained.

Numerically, the least-squares problem is best solved by singular value decomposition of $A\Delta = \mathbf{b}$, since the formation of AA^T significantly diminishes the obtainable precision. Nevertheless, the method has difficulties with flat cusps — for example, with singularities such as $y^2 - x^{2m+1} = 0$, $m \gg 1$.

Using Constrained Minimization

We may consider locating a singularity as a constrained-minimization problem:

Problem
Minimize $f_x^2 + f_y^2$ subject to the constraint $f(x, y) = 0$.

With the help of Lagrangian multipliers, this problem can be converted to an unconstrained minimization problem by minimizing

$$L = f_x^2 + f_y^2 + \lambda f$$

where λ is the Lagrange multiplier. An extremum of L then satisfies the following equations:

$$\begin{aligned} L_x &= 0 \\ L_y &= 0 \\ L_\lambda &= 0 \end{aligned} \tag{6.15}$$

These are three algebraic equations in three variables; that is, they represent the intersection of three algebraic surfaces in (x, y, λ)-space. Given an initial guess (x_0, y_0), the intersection will include a nearby curve singularity. An initial guess for λ must also be given; it could be 1, for instance.

Newton's method cannot be used to solve the system of equations (6.15) without attention to some details. To understand the reason, we expand the partials of the goal function L from which the matrix for the Newton iteration is formed.

$$\begin{aligned} L_x &= 2(f_x f_{xx} + f_y f_{xy}) + \lambda f_x \\ L_y &= 2(f_x f_{xy} + f_y f_{yy}) + \lambda f_y \\ L_\lambda &= f \end{aligned}$$

The matrix of the linear system used in Newton's method is therefore

$$\begin{pmatrix} u_{11} & u_{12} & f_x \\ u_{21} & u_{22} & f_y \\ f_x & f_y & 0 \end{pmatrix} \tag{6.16}$$

where

$$\begin{aligned} u_{11} &= 2(f_{xx}^2 + f_x f_{xxx} + f_{xy}^2 + f_y f_{xxy}) + \lambda f_{xx} \\ u_{12} &= 2(f_{xx} f_{xy} + f_x f_{xxy} + f_{xy} f_{yy} + f_y f_{xyy}) + \lambda f_{xy} \\ u_{21} &= 2(f_{xx} f_{xy} + f_x f_{xxy} + f_{xy} f_{yy} + f_y f_{xyy}) + \lambda f_{xy} \\ u_{22} &= 2(f_{xy}^2 + f_x f_{xyy} + f_{yy}^2 + f_y f_{yyy}) + \lambda f_{yy} \end{aligned}$$

Since $f_x \to 0$ and $f_y \to 0$ as we approach the singularity, the last row vanishes and matrix (6.16) becomes singular. However, the last row governs the change in λ, which is irrelevant as long as we stay on the curve f. So, we could restrict attention to the 2×2 submatrix

$$\begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \tag{6.17}$$

This modification will not help for higher-order singularities where, in addition to f_x and f_y, the higher-order partials f_{xx}, f_{xy}, and f_{yy} vanish as well.

Reduction to Root Finding

Locating the singularities of the curve $f(x, y) = 0$ can be reduced to finding the roots of a univariate polynomial as follows. By forming the resultant $Res_y(f, f_x)$, we obtain a univariate polynomial $P(x)$. Its roots are the x coordinates of those curve points at which f_x vanishes. The partial derivative f_x vanishes at all singular curve points, and it vanishes at all regular curve points at which the curve tangent is parallel to the x axis. So, we must determine $P(x)$, and, for each root a, we need to test whether there is a singular curve point (a, y_a). The ordinates y_a can be determined as the roots of the polynomial $f(a, y)$.

Example 6.8: Consider the curve

$$f = x^3 + 3x^2y + 3xy^2 + y^3 - 3x^2 - 2xy - 3y^2 + 5x - 3y - 4$$

The partial derivatives are

$$\begin{aligned} f_x &= 3x^2 + 6xy + 3y^2 - 6x - 2y + 5 \\ f_y &= 3x^2 + 6xy + 3y^2 - 2x - 6y - 3 \end{aligned}$$

The y-resultant of f and f_x is

$$Res_y(f, f_x) = 432x^4 - 2464x^3 + 5256x^2 - 4968x + 1755$$

with the two real roots

$$\begin{aligned} a_1 &= 1.5 \\ a_2 &= 1.2037037... \end{aligned}$$

For the abscissa a_1 the only ordinate value is $b_1 = -0.5$. For the other real root, we obtain the ordinate values $b_{21} = -1.0925926$, $b_{22} = -0.5290596$, and $b_{23} = -0.5290596$. Here, f_x does not vanish at (a_2, b_{22}) and at (a_2, b_{23}), and f_y does not vanish at (a_2, b_{21}); hence, all three points are regular. However, both f_x and f_y vanish at (a_1, b_1), which therefore is a singularity of f. $\diamond$

From a practical perspective, the cost of computing $P(x)$ can be expected to be noticeable. On the other hand, the computation needs to be done only once. A second reduction method to root finding, based on Gröbner bases techniques, will be discussed in Section 7.4.3 of Chapter 7.

6.5.6 Bringing the Singularity to the Origin

When we are determining the locus of a nearby singularity numerically, the resulting coordinates will be imprecise. However, the validity of the transformations T_1 and T_2 depends on the origin being a singularity. Intuitively, small perturbations in the position of the singularity will introduce low-order terms with small coefficients in the translated polynomial $\bar{f}$. These terms should be eliminated. To identify them, we recall that the term cx^iy^j occurs in f iff the partial derivative

$$\frac{1}{i!j!}\,\frac{\partial^{i+j}}{\partial x^i \partial y^j} f(0,0) \;=\; c$$

Since the value of partial derivatives is invariant under translation, it follows that the term cx^iy^j should be absent in $\bar{f}$ iff the corresponding partial vanishes at the singularity located prior to translation. Vanishing low-order partials are discovered as part of the least-squares iteration, but must be determined explicitly in the case of other methods for determining the singularity.

6.5.7 Preserving the Direction of Tracing

At nonsingular points, we locally orient f by the tangent vector $(-f_y, f_x)$. At a singularity, curve segments locally belonging to the same analytic branch may be oriented in an opposite direction, necessitating a reversal of the nominal tracing direction. An example is shown in Figure 6.17. We establish a relationship between the orientation of the curve f and the orientation of its proper transform g with the goal of recognizing when to reverse the tracing direction after having passed through a singular point.

Let $p = (a_0, b_0)$ be a nonsingular point of f, where $a_0 \neq 0$. Let

$$\begin{aligned} x(s) &= a_0 + a_1 s + a_2 s^2 + \ldots \\ y(s) &= b_0 + b_1 s + b_2 s^2 + \ldots \end{aligned}$$

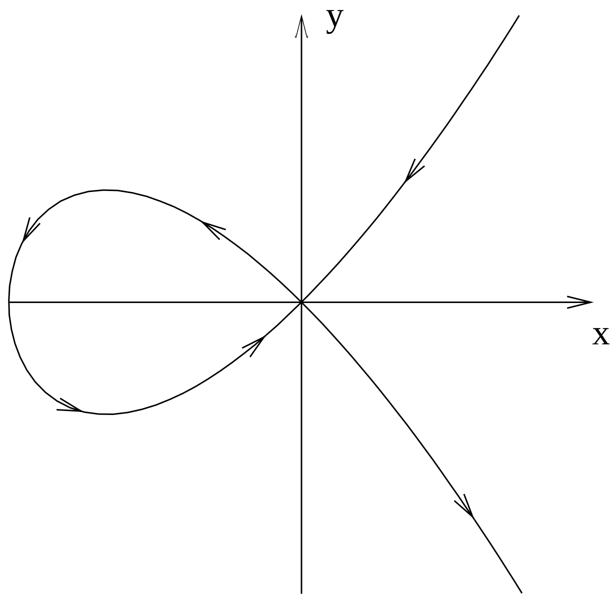


Figure 6.17 Curve Orientation of $f = y^2 - x^2 - x^3$ by $(-f_y, f_x)$

be the place of f centered at p. The place defines a branch orientation by increasing s. This orientation agrees with the standard orientation $(-f_y, f_x)$ whenever

$$\begin{aligned}\operatorname{sign}(f_x(a_0, b_0)) &= \operatorname{sign}(b_1)\\ \operatorname{sign}(f_y(a_0, b_0)) &= \operatorname{sign}(-a_1)\end{aligned}$$

Otherwise, it is opposite. At the corresponding point $p_1 = (a_0, b_0/a_0)$ of g, the transformed curve g has the place

$$\begin{aligned} x_1(s) &= x(s) \\ y_1(s) &= y(s)/x(s) = c_0 + c_1 s + c_2 s^2 + \ldots \end{aligned}$$

Since $x(s) = x_1(s)$, the curve and its transform are oriented the same way, by increasing s. Moreover, dividing $y(s)$ by $x(s)$, we obtain

$$\begin{aligned} c_0 &= b_0/a_0 \\ c_1 &= (b_1 a_0 - a_1 b_0)/a_0^2 \end{aligned}$$

and so on. Hence, relating this to the standard orientation by a proportionality constant α, we have

$$\begin{aligned} g_y &= \alpha f_y \\ g_x &= \alpha(x f_x + y f_y)/x^2 \end{aligned}$$

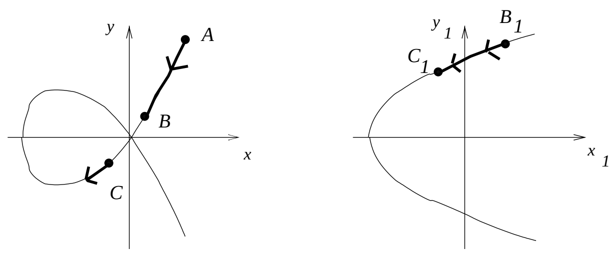


Figure 6.18 Correlation of Standard Curve Orientations

If sign(α) = 1, then both f and g are traced in the same direction, relative to the standard orientation; otherwise, we trace g in the opposite direction. Since no orientation reversal can happen on the fully desingularized branch, we obtain the following recursive algorithm for maintaining a consistent direction of traversal:

1. Traverse f in the direction $u(-f_y, f_x)$, where $u = 1$ or $u = -1$.

2. When changing over to the proper transform g of f, compute the sign of α for the corresponding points p of f and p_1 of g at which we switch.

3. If $\alpha > 0$, the transform g is traversed in the direction $u(-g_y, g_x)$; otherwise, it is traversed in the opposite direction.

4. Assume recursively that we have the traversal direction $u'(-g_y, g_x)$ at a regular point p'_1 of g. When reverting to tracing f at the corresponding point p', compute the sign of α again, and, if necessary, complement u.

Example 6.9: Figure 6.18 illustrates the method for maintaining consistent traversal direction in the presence of singularities. Assume we are tracing the curve $f : y^2 - x^2 - x^3 = 0$ beginning at the point A in the direction $(-f_y, f_x)$. At point B, we switch to tracing the proper transform $g : y_1^2 - 1 - x_1 = 0$ because of the impending singularity at the origin. We determine the proportionality factor between the partials of f at B and the partials of g at the corresponding point B_1. We find that α is positive, and trace g in the direction $(-g_y, g_x)$. At the point C_1 of g, we determine that we have safely passed the singularity and switch back to tracing f at the corresponding point C. Again, we investigate α to find that this time it is negative. Therefore, we continue traversing f in the direction $(f_y, -f_x)$. $\diamond$

6.6 Remarks on Surface Intersection

Sections 6.4 and 6.5 showed how to reduce surface intersection to the evaluation of a plane algebraic curve. As mentioned before, the ability of the plane-curve tracing method to cope with complicated singularities is a strong advantage of this approach, especially since, until now, a comprehensive method for handling surface-intersection singularities has not been developed and implemented. However, the approach has to cope with a number of practical difficulties, including the following:

1. Implicitizing a parametrically defined surface entails substantial symbolic computation for surfaces such as bicubic patches.

2. Although it is conceptually trivial to substitute a parametric into an implicit form, it is subtle in practice, because of possible floating-point errors. Exact arithmetic is, of course, free from such errors, but is more expensive.

3. The plane algebraic curve eventually obtained is typically of very high algebraic degree. For example, the intersection of two bicubic surface is in general an algebraic curve of degree 324. At such high degrees, severe numerical problems are often encountered.

Many of these difficulties are remedied by tracing in higher dimensions. For instance, the intersection of two bicubics is easily traced in four dimensions, and ordinary double-precision floating-point arithmetic delivers accurate results. On the other hand, higher-dimensional approaches are traditionally perceived as being slow. Probably, this perception can be corrected with a sophisticated implementation of the method. However, the method raises a number of research issues to which there are, at this time, only partial answers. We mention the following issues:

1. Finding a starting point for the trace is harder in higher dimensions. It is certain that subdivision and domain-shrinking techniques can be generalized directly; however, unless care is exercised, the complexity of these methods grows exponentially with the dimension.

2. In a higher-dimensional space, complicated singularities might be present on the intersection curve. These could be difficult to analyze and resolve.

3. By elimination, it is always possible to reduce the number of equations and variables at the expense of raising the degree of the algebraic equations. The tradeoff entailed by this strategy is not understood, and no method is known for introducing additional variables and equations that lowers the algebraic degrees of the equations.

More research and experimentation is needed to understand these issues in depth, and to assess the proper role of the different approaches to surface intersection.

6.7 Notes and References

In many cases, initial starting points for the trace will be intersections of the curve to be traced with another surface. In that case, the methods from Section 7.4 in Chapter 7 could be used. There are other methods, including subdivision strategies, lattice methods, and homotopy continuation. See, for example, Pratt and Geisow (1986) and Morgan (1987).

The material of Section 6.2 is mostly from Bajaj, Hoffmann, Hopcroft, and Lynch (1988). Our exposition of singular value decomposition is based on Golub and van Loan (1983). For an elementary introduction to differential geometry and the properties of the moving triad, see Hilbert and Cohn-Vossen (1952). Montaudouin, Tiller, and Vold (1986) discuss step-size selection in the context of an explicit approximant to a plane curve; that is, they use an approximant of the form $y = a_0 + a_1 x + a_2 x^2 + \cdots$, based on the implicit function theorem.

The idea of tracing surface intersections in higher dimensions is mentioned in Bajaj, Hoffmann, Hopcroft, and Lynch (1988), where it is applied to the numerical intersection of parametric surfaces. Hoffmann (1988) discusses applications of the higher-dimensional formulation, including offsets, equal-distance surfaces, and variable-radius blending surfaces. The paper explores the suitability of using the higher-dimensional formulation as a representation for such surfaces, and reports experiences with an implementation. Subdivision and approximation methods for surfaces defined as $n - 2$ algebraic equations in n variables are currently being explored.

Farouki and Neff (1989) discuss the algebraic formulation of offsets of rational plane curves, and show that the extraneous points at singularities can be eliminated by dividing out common factors of the derivatives of the coordinate functions. It is unclear whether this approach generalizes to implicit algebraic curves. The envelope theorem is discussed in Spivak (1975), Volume 2. Farouki and Neff (1989) contains figures similar to our figures 6.6 and 6.7.

There is a rich literature on quadratic transformations and on the desingularization of plane algebraic curves. For a projective version, see, for example, van der Waerden (1939). For an affine version, see, for example, Abhyankar (1983). Semple and Kneebone (1959) give a lucid description of the effect of quadratic transformations on the curve places at a singularity. Roughly speaking, the effect of quadratic transformations on the power series is akin to a shifting operation in which the coefficients a_0 and b_0 are dropped, and the remaining coefficients shift to the left. Since each place has a regular structure for sufficiently high exponents of s, repeated quadratic transformations eventually strip off the leading, irregular coefficient structures, leaving only the trailing, regular part. When that has happened, the place has been desingularized, and the various places can be separated. See Walker (1950) for a discussion of places, order of contact of places, tangents to places, and other properties.

The formulation of nonlinear equations for determining the places at a singularity, and the correspondence with the system of equations (6.1), is from Hoffmann (1988), who also notes the monoid computation when seeking to map the intersection of implicit surfaces to plane curves. The monoid computation was well known in the nineteenth-century geometry literature, where it was used as a technical device for classifying different types of space curves of low degree. See, for example, Snyder and Sisam (1914). Owen and Rockwood (1987) analyze singularities on surface/surface intersections by constructing locally a second-order approximant to each surface and analyzing the possible types of singularities of the approximants. The method cannot deal with all types of singularities.

Mapping surface intersections to plane algebraic curves has been advocated repeatedly — by Geisow (1983), by Farouki (1986a), and by many other authors. The approach has many appealing aspects, but so far does not seem to have had a deep effect on practice. Because substitution is such a natural conceptual operation, some authors overlook that it may introduce substantial floating-point errors. A good discussion of this and related practical points can be found in Prakash and Patrikalakis (1988).

The section on projection methods presents material from Abhyankar and Bajaj (1987d). Garrity and Warren (1988) propose a different method for projecting implicit surface intersections to plane algebraic curves. Their method removes the requirement that the two surfaces must intersect transversally in an irreducible space curve, and is based on a classical theorem that says that all but finitely many points on a line are good projection points, provided the line does not intersect the space curve.

Locating singularities precisely is one of the main practical concerns when implementing the plane curve trace. The least-squares approach is from Bajaj, Hoffmann, Hopcroft, and Lynch (1988). Its convergence can be slow, but the use of overrelaxation could improve the convergence rate. Using constrained optimization is a natural idea. Prakash and Patrikalakis (1988) use the technique with good success for ordinary singularities. In the case of higher-order singularities, they suggest an adaptive extension of the goal function to be minimized. The extension uses higher-order partial derivatives in a manner similar to the least-squares method described. Sederberg (1988) proposes a subdivision scheme for locating singularities that assumes that the curve has been expressed in a special Bernstein basis form. For a method to translate a polynomial into Bernstein form see, for example, Waggenspack and Anderson (1986).