Chapter 1

Introduction

Computer aided geometric design (CAGD) concerns itself with the mathematical description of shape for use in computer graphics, manufacturing, or analysis. It draws upon the fields of geometry, computer graphics, numerical analysis, approximation theory, data structures and computer algebra.

CAGD is a young field. The first work in this field began in the mid 1960s. The term *computer aided geometric design* was coined in 1974 by R.E. Barnhill and R.F. Riesenfeld in connection with a conference at the University of Utah.

This chapter presents some basic background material such as vector algebra, equations for lines and conic sections, homogeneous coordinates.

1.1 Points, Vectors and Coordinate Systems

Consider the simple problem of writing a computer program which finds the area of any triangle. We must first decide how to uniquely describe the triangle. One way might be to provide the lengths l_1, l_2, l_3 of the three sides, from which Heron's formula yields

$$Area = \sqrt{s(s-l_1)(s-l_2)(s-l_3)}, \qquad s = \frac{l_1+l_2+l_3}{2}.$$

An alternative way to describe the triangle is in terms of its vertices. But while the lengths of the sides of a triangle are independent of its position, we can specify the vertices to our computer program only with reference to some *coordinate system* — which can be defined simply as any method for representing points with numbers.

Note that a coordinate system is an artificial devise which we arbitrarily impose for the purposes at hand. Imagine a triangle cut out of paper and lying on a flat table in the middle of a room. We could define a *Cartesian coordinate system* whose origin lies in a corner of the room, and whose coordinate axes lie along the three room edges which meet at the corner. We would further specify the unit of measurement, say centimeters. Then, each vertex of our triangle could be described in terms of its respective distance from the two walls containing the origin and from the floor. These distances are the Cartesian coordinates (x, y, z) of the vertex with respect to the coordinate system we defined.

Vectors A *vector* can be pictured as a line segment of definite length with an arrow on one end. We will call the end with the arrow the *tip* or *head* and the other end the *tail*.

Two vectors are equivalent if they have the same length, are parallel, and point in the same direction (have the same *sense*) as shown in Figure 1.1. For a given coordinate system, we can

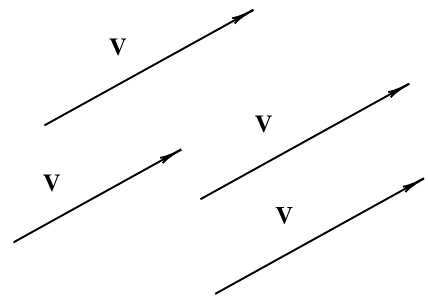
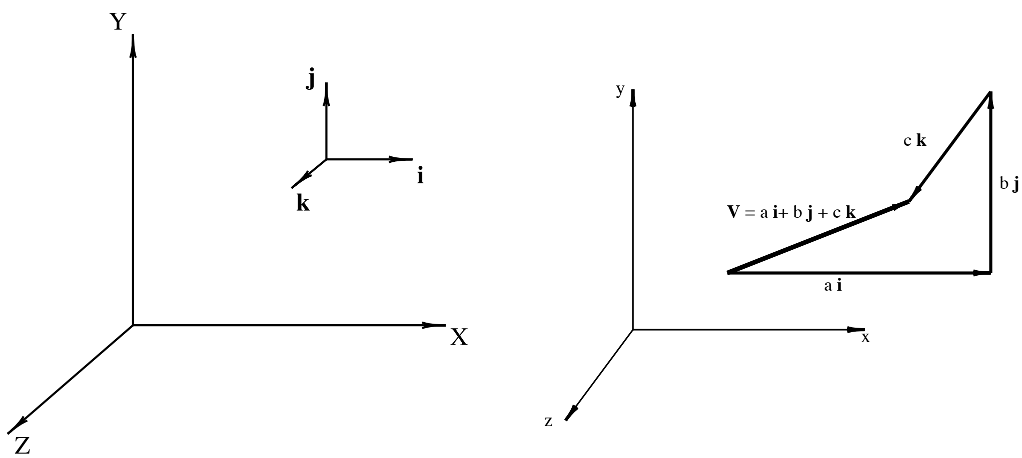


Figure 1.1: Equivalent Vectors

describe a three-dimensional vector in the form (a, b, c) where a (or b or c) is the distance in the x (or y or z) direction from the tail to the tip of the vector.



(a) Unit Vectors. (b) Vector in Component Form.

Figure 1.2: Vectors.

Unit Vectors The symbols **i**, **j**, and **k** denote vectors of "unit length" (based on the unit of measurement of the coordinate system) which point in the positive x, y, and z directions respectively (see Figure 1.2.a).

Unit vectors allow us to express a vector in component form (see Figure 1.2.b):

$$\mathbf{P} = (a, b, c) = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

An expression such as (x, y, z) can be called a *triple* of numbers. In general, an expression $(x_1, x_2, \ldots, x_n)$ is an n-tuple, or simply a tuple. As we have seen, a triple can signify either a point or a vector.

Relative Position Vectors Given two points $\mathbf{P}_1$ and $\mathbf{P}_2$, we can define

$$\mathbf{P}_{2/1} = \mathbf{P}_2 - \mathbf{P}_1$$

as the vector pointing from $\mathbf{P}_1$ to $\mathbf{P}_2$. This notation $\mathbf{P}_{2/1}$ is widely used in engineering mechanics, and can be read "the position of point $\mathbf{P}_2$ relative to $\mathbf{P}_1$"

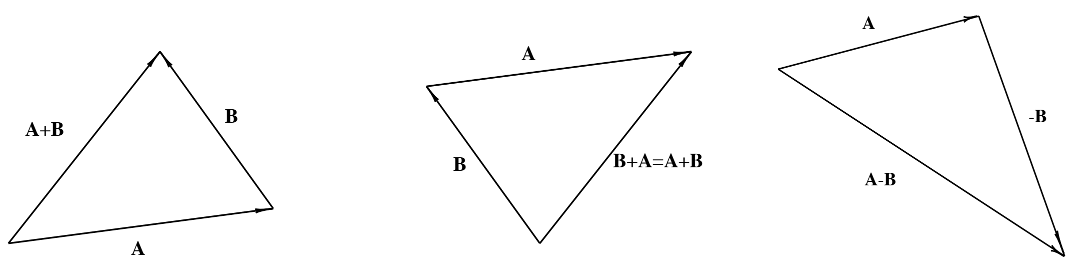
In our diagrams, points will be drawn simply as dots or small circles, and vectors as line segments with single arrows. Vectors and points will both be denoted by bold faced type.

1.2 Vector Algebra

Given two vectors $\mathbf{P}_1 = (x_1, y_1, z_1)$ and $\mathbf{P}_2 = (x_2, y_2, z_2)$, the following operations are defined:

Addition:

$$\mathbf{P}_1 + \mathbf{P}_2 = \mathbf{P}_2 + \mathbf{P}_1 = (x_1 + x_2, y_1 + y_2, z_1 + z_2)$$



(a) Vector Addition. (b) Vector Subtraction.

Figure 1.3: Vector Addition and Subtraction.

Subtraction:

$$\mathbf{P}_1 - \mathbf{P}_2 = (x_1 - x_2, y_1 - y_2, z_1 - z_2)$$

Scalar multiplication:

$$c\mathbf{P}_1 = (cx_1, cy_1, cz_1)$$

Length of a Vector

$$|\mathbf{P}_1| = \sqrt{x_1^2 + y_1^2 + z_1^2}$$

A vector of length one is called a unit vector.

$$\frac{\mathbf{P}_1}{|\mathbf{P}_1|}$$

is a unit vector in the direction of $\mathbf{P}_1$

Dot Product The dot product of two vectors is defined

$$\mathbf{P}_1 \cdot \mathbf{P}_2 = |\mathbf{P}_1||\mathbf{P}_2| \cos\theta \tag{1.1}$$

where θ is the angle between the two vectors. Since the unit vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are mutually perpendicular,

$$\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1$$

$$\mathbf{i} \cdot \mathbf{j} = \mathbf{i} \cdot \mathbf{k} = \mathbf{j} \cdot \mathbf{k} = 0.$$

The dot product obeys the distributive law

$$\mathbf{P}_1 \cdot (\mathbf{P}_2 + \mathbf{P}_3) = \mathbf{P}_1 \cdot \mathbf{P}_2 + \mathbf{P}_1 \cdot \mathbf{P}_3,$$

As a result of the distributive law,

$$\begin{aligned} \mathbf{P}_1 \cdot \mathbf{P}_2 &= (x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}) \cdot (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) \\ &= (x_1 * x_2 + y_1 * y_2 + z_1 * z_2) \end{aligned} \tag{1.2}$$

(1.2) enables us to compute the angle between any two vectors. From (1.1),

$$\theta = \cos^{-1}\left(\frac{\mathbf{P}_1 \cdot \mathbf{P}_2}{|\mathbf{P}_1||\mathbf{P}_2|}\right).$$

Example. Find the angle between vectors $(1, 2, 4)$ and $(3, -4, 2)$.

Answer.

$$\begin{aligned} \theta &= \cos^{-1}\left(\frac{\mathbf{P}_1 \cdot \mathbf{P}_2}{|\mathbf{P}_1||\mathbf{P}_2|}\right) \\ &= \cos^{-1}\left(\frac{(1,2,4)\cdot(3,-4,2)}{|(1,2,4)||(3,-4,2)|}\right) \\ &= \cos^{-1}\left(\frac{3}{\sqrt{21}\sqrt{29}}\right) \\ &\approx 83.02^\circ \end{aligned}$$

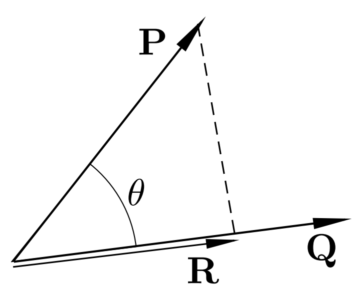


Figure 1.4: Vector Projection

An important application of dot products is in computing the projection of one vector onto another vector. As illustrated in Figure 1.4, vector $\mathbf{R}$ is the projection of vector $\mathbf{P}$ onto vector $\mathbf{Q}$. Since

$$|\mathbf{R}| = |\mathbf{P}| \cos\theta = \frac{\mathbf{P} \cdot \mathbf{Q}}{|\mathbf{Q}|}$$

we have

$$\mathbf{R} = |\mathbf{R}|\frac{\mathbf{Q}}{|\mathbf{Q}|} = \frac{\mathbf{P}\cdot\mathbf{Q}}{|\mathbf{Q}|}\frac{\mathbf{Q}}{|\mathbf{Q}|} = \frac{\mathbf{P}\cdot\mathbf{Q}}{\mathbf{Q}\cdot\mathbf{Q}}\mathbf{Q}.$$

Example. Find the projection of $\mathbf{P} = (3, 2, 1)$ onto $\mathbf{Q} = (3, 6, 6)$.

Answer.

$$\begin{aligned} \mathbf{R} &= \frac{\mathbf{P}\cdot\mathbf{Q}}{\mathbf{Q}\cdot\mathbf{Q}}\mathbf{Q} \\ &= \frac{(3,2,1)\cdot(3,6,6)}{(3,6,6)\cdot(3,6,6)}(3,6,6) \\ &= \frac{27}{81}(3,6,6) \\ &= (1,2,2) \end{aligned}$$

Cross Product: The cross product $\mathbf{P}_1 \times \mathbf{P}_2$ is a vector whose magnitude is

$$|\mathbf{P}_1 \times \mathbf{P}_2| = |\mathbf{P}_1||\mathbf{P}_2|\sin\theta$$

(where again θ is the angle between $\mathbf{P}_1$ and $\mathbf{P}_2$), and whose direction is mutually perpendicular to $\mathbf{P}_1$ and $\mathbf{P}_2$ with a sense defined by the right hand rule as follows. Point your fingers in the direction of $\mathbf{P}_1$ and orient your hand such that when you close your fist your fingers pass through the direction of $\mathbf{P}_2$. Then your right thumb points in the sense of $\mathbf{P}_1 \times \mathbf{P}_2$.

From this basic definition, one can verify that

$$\mathbf{P}_1 \times \mathbf{P}_2 = -\mathbf{P}_2 \times \mathbf{P}_1,$$

$$\mathbf{i}\times\mathbf{j} = \mathbf{k}, \quad \mathbf{j}\times\mathbf{k} = \mathbf{i}, \quad \mathbf{k}\times\mathbf{i} = \mathbf{j}$$

$$\mathbf{j}\times\mathbf{i} = -\mathbf{k}, \quad \mathbf{k}\times\mathbf{j} = -\mathbf{i}, \quad \mathbf{i}\times\mathbf{k} = -\mathbf{j}.$$

The cross product obeys the distributive law

$$\mathbf{P}_1 \times (\mathbf{P}_2 + \mathbf{P}_3) = \mathbf{P}_1 \times \mathbf{P}_2 + \mathbf{P}_1 \times \mathbf{P}_3.$$

This leads to the important relation

$$\begin{aligned} \mathbf{P}_1 \times \mathbf{P}_2 &= (x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}) \times (x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}) \\ &= (y_1z_2 - y_2z_1, x_2z_1 - x_1z_2, x_1y_2 - x_2y_1) \\ &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix} \end{aligned} \tag{1.3}$$

Area of a Triangle. Cross products have many important uses, such as finding a vector which is mutually perpendicular to two other vectors and finding the area of a triangle which is defined by three points $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$.

$$Area = \frac{1}{2}|\mathbf{P}_{1/2}||\mathbf{P}_{1/3}|\sin\theta_1 = \frac{1}{2}|\mathbf{P}_{1/2} \times \mathbf{P}_{1/3}| \tag{1.4}$$

For example, the area of a triangle with vertices $\mathbf{P}_1 = (1,1,1)$, $\mathbf{P}_2 = (2,4,5)$, $\mathbf{P}_3 = (3,2,6)$ is

$$\begin{aligned} Area &= \frac{1}{2}|\mathbf{P}_{1/2} \times \mathbf{P}_{1/3}| \\ &= \frac{1}{2}|(1,3,4) \times (2,1,5)| \\ &= \frac{1}{2}|(11,3,-5)| = \frac{1}{2}\sqrt{11^2 + 3^2 + (-5)^2} \\ &\approx 6.225 \end{aligned}$$

1.2.1 Points vs. Vectors

A point is a geometric entity which connotes position, whereas a vector connotes direction and magnitude. From a purely mathematical viewpoint, there are good reasons for carefully distinguishing between triples that refer to points and triples that signify vectors [Goldman '85]. However, no problem arises if we recognize that a triple connoting a point can be interpreted as a vector from the origin to the point. Thus, we could call a point an *absolute position vector* and the difference between two points a *relative position vector*. These phrases are often used in engineering mechanics, where vectors are used to express quantities other than position, such as velocity or acceleration.

1.3 Rotation About an Arbitrary Axis

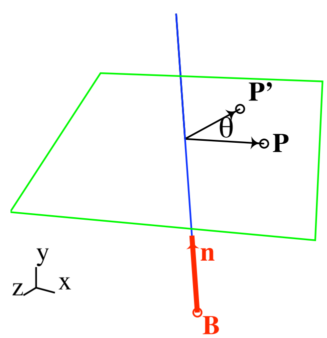
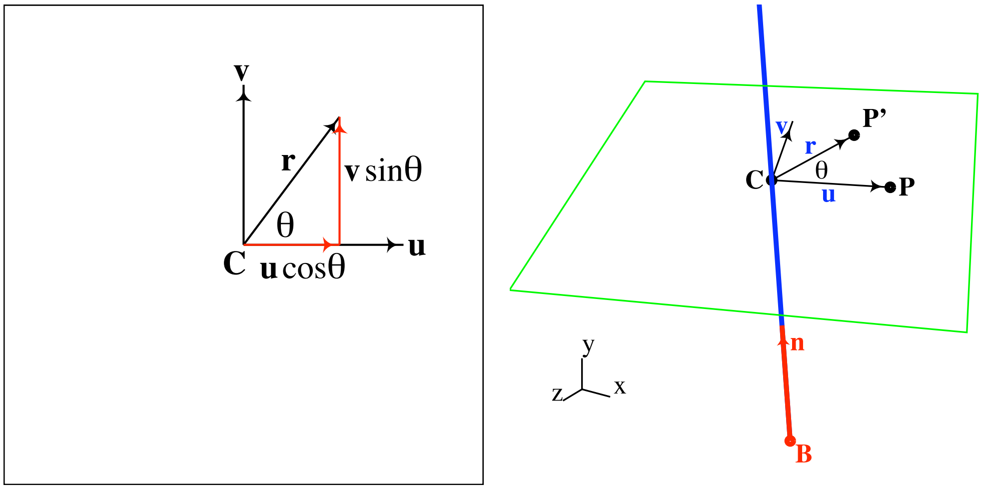


Figure 1.5: Rotation about an Arbitrary Axis

The problem of computing a rotation about an arbitrary axis is fundamental to CAGD and computer graphics. The standard solution to this problem as presented in most textbooks on computer graphics involves the concatenation of seven 4×4 matrices. We present here a straightforward solution comprised of the four simple vector computations in equations (1.6) through (1.9) — a compelling example of the power of vector algebra.

Figure 1.5 shows a point $\mathbf{P}$ which we want to rotate an angle θ about an axis that passes through $\mathbf{B}$ with a direction defined by unit vector $\mathbf{n}$. So, given the angle θ, the unit vector $\mathbf{n}$, and Cartesian coordinates for the points $\mathbf{P}$ and $\mathbf{B}$, we want to find Cartesian coordinates for the point $\mathbf{P}'$.

The key insight needed is shown in Figure 1.6.a.



(a) Key Insight. (b) Labels.

Figure 1.6: Rotation about an Arbitrary Axis Using Vector Algebra.

Let $\mathbf{u}$ and $\mathbf{v}$ be any two three-dimensional vectors that satisfy $u \cdot v = 0$ (that is, they are perpendicular) and $|\mathbf{u}| = |\mathbf{v}| \neq 0$ (that is, they are they same length but not necessarily unit vectors). We want to find a vector $\mathbf{r}$ that is obtained by rotating $\mathbf{u}$ an angle θ in the plane defined by $\mathbf{u}$ and $\mathbf{v}$. As suggested in Figure 1.6,

$$\mathbf{r} = \mathbf{u}\cos\theta + \mathbf{v}\sin\theta. \tag{1.5}$$

With that insight, it is easy to compute a rotation about an arbitrary axis. Note that $(\mathbf{C} - \mathbf{B})$ is the projection of vector $(\mathbf{P} - \mathbf{B})$ onto the unit vector $\mathbf{n}$. Referring to the labels in Figure 1.6.b, we compute

$$\mathbf{C} = \mathbf{B} + [(\mathbf{P} - \mathbf{B}) \cdot \mathbf{n}]\mathbf{n}. \tag{1.6}$$

$$\mathbf{u} = \mathbf{P} - \mathbf{C} \tag{1.7}$$

$$\mathbf{v} = \mathbf{n} \times \mathbf{u} \tag{1.8}$$

Then, $\mathbf{r}$ is computed using equation (1.5), and

$$\mathbf{P}' = \mathbf{C} + \mathbf{r}. \tag{1.9}$$

Example
Find the coordinates of a point $(5, 7, 3)$ after it is rotated an angle $\theta = 90°$ about an axis that points in the direction $(2, 1, 2)$ and that passes through the point $(1, 1, 1)$.
Answer:
$\mathbf{n} = (\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$.

$$\begin{aligned} \mathbf{C} &= \mathbf{B} + [(\mathbf{P} - \mathbf{B}) \cdot \mathbf{n}]\mathbf{n} \\ &= (1,1,1) + [((5,7,3) - (1,1,1)) \cdot (\frac{2}{3}, \frac{1}{3}, \frac{2}{3})](\frac{2}{3}, \frac{1}{3}, \frac{2}{3}) \\ &= (5,3,5) \end{aligned} \tag{1.10}$$

$$\mathbf{u} = \mathbf{P} - \mathbf{C} = (0, 4, -2)$$

$$\mathbf{v} = \mathbf{n} \times \mathbf{u} = (-\frac{10}{3}, \frac{4}{3}, \frac{8}{3}).$$

$$\mathbf{r} = \mathbf{u}\cos\theta + \mathbf{v}\sin\theta = \mathbf{u}\cdot 0 + \mathbf{v}\cdot 1 = \mathbf{v}.$$

1.3.1 Matrix Form

It is possible to take these simple vector equations and to create from them a single 4×4 transformation matrix for rotation about an arbitrary axis. While this is useful to do in computer graphics (where, in fact, this matrix is typically created by concatenating seven 4×4 matrices), the simple vector equations we just derived suffice for many applications. The derivation of this matrix is presented here for your possible reference. We will not be using it in this course.

Let $\mathbf{P} = (x, y, z)$, $\mathbf{P}' = (x', y', z')$, $\mathbf{B} = (B_x, B_y, B_z)$, and $\mathbf{n} = (n_x, n_y, n_z)$. We seek a 4×4 matrix M such that

$$M \left\{ \begin{array}{c} x \\ y \\ z \\ 1 \end{array} \right\} = \left\{ \begin{array}{c} x' \\ y' \\ z' \\ 1 \end{array} \right\}$$

$$(C_x, C_y, C_z) = (B_x, B_y, B_z) + [xn_x + yn_y + zn_z - \mathbf{B}\cdot\mathbf{n}](n_x, n_y, n_z) \tag{1.11}$$

$$\begin{aligned} C_x &= xn_x^2 + yn_xn_y + zn_xn_z + B_x - (\mathbf{B}\cdot\mathbf{n})n_x && (1.12)\\ C_y &= xn_xn_y + yn_y^2 + zn_yn_z + B_y - (\mathbf{B}\cdot\mathbf{n})n_y && (1.13)\\ C_z &= xn_xn_z + yn_yn_z + zn_z^2 + B_z - (\mathbf{B}\cdot\mathbf{n})n_z && (1.14) \end{aligned}$$

$$\mathbf{u} = (x, y, z) - (C_x, C_y, C_z) \tag{1.15}$$

$$\begin{aligned} u_x &= x(1 - n_x^2) - yn_xn_y - zn_xn_z + (\mathbf{B}\cdot\mathbf{n})n_x - B_x && (1.16)\\ u_y &= -xn_xn_y + y(1 - n_y^2) - zn_yn_z + (\mathbf{B}\cdot\mathbf{n})n_y - B_y && (1.17)\\ u_z &= -xn_xn_z - yn_yn_z + z(1 - n_z^2) + (\mathbf{B}\cdot\mathbf{n})n_z - B_z && (1.18) \end{aligned}$$

$$\begin{aligned} v_x &= n_yu_z - n_zu_y && (1.19)\\ v_y &= n_zu_x - n_xu_z && (1.20)\\ v_z &= n_xu_y - n_yu_x && (1.21) \end{aligned}$$

$$\begin{aligned} r_x &= u_x\cos\theta + (n_yu_z - n_zu_y)\sin\theta && (1.22)\\ r_y &= u_y\cos\theta + (n_zu_x - n_xu_z)\sin\theta && (1.23)\\ r_z &= u_z\cos\theta + (n_xu_y - n_yu_x)\sin\theta && (1.24) \end{aligned}$$

$$(x', y', z') = (C_x + r_x, C_y + r_y, C_z + r_z) \tag{1.25}$$

$$\begin{aligned} x' \quad &= xn_x^2 + yn_xn_y + zn_xn_z + B_x - (\mathbf{B}\cdot\mathbf{n})n_x + \\ &[x(1-n_x^2) - yn_xn_y - zn_xn_z + (\mathbf{B}\cdot\mathbf{n})n_x - B_x]\cos\theta + \\ &n_y[-xn_xn_z - yn_yn_z + z(1-n_z^2) + (\mathbf{B}\cdot\mathbf{n})n_z - B_z]\sin\theta - \\ &n_z[-xn_xn_y + y(1-n_y^2) - zn_yn_z + (\mathbf{B}\cdot\mathbf{n})n_y - B_y]\sin\theta \end{aligned} \tag{1.26}$$

$$\begin{aligned} x' &= x[n_x^2 + (1-n_x^2)\cos\theta] + y[n_xn_y(1-\cos\theta) - n_z\sin\theta] \\ &+ z[n_xn_z(1-\cos\theta) + n_y\sin\theta] + (B_x - (\mathbf{B}\cdot\mathbf{n})n_x)(1-\cos\theta) + (n_zB_y - n_yB_z)\sin\theta. \end{aligned}$$

Since $n_x^2 + n_y^2 + n_z^2 = 1$, $(1-n_x^2) = n_y^2 + n_z^2$. In like manner we can come up with an expression for y' and z', and our matrix M is thus

$$\begin{bmatrix} n_x^2 + (n_y^2+n_z^2)\cos\theta & n_xn_y(1-\cos\theta) - n_z\sin\theta & n_xn_z(1-\cos\theta) + n_y\sin\theta & T_1 \\ n_xn_y(1-\cos\theta) + n_z\sin\theta & n_y^2 + (n_x^2+n_z^2)\cos\theta & n_yn_z(1-\cos\theta) - n_x\sin\theta & T_2 \\ n_xn_z(1-\cos\theta) - n_y\sin\theta & n_yn_z(1-\cos\theta) + n_x\sin\theta & n_z^2 + (n_x^2+n_y^2)\cos\theta & T_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1.27}$$

with

$$T_1 = (B_x - (\mathbf{B}\cdot\mathbf{n})n_x)(1-\cos\theta) + (n_zB_y - n_yB_z)\sin\theta$$
$$T_2 = (B_y - (\mathbf{B}\cdot\mathbf{n})n_y)(1-\cos\theta) + (n_xB_z - n_zB_x)\sin\theta$$
$$T_3 = (B_z - (\mathbf{B}\cdot\mathbf{n})n_z)(1-\cos\theta) + (n_yB_x - n_xB_y)\sin\theta$$

1.4 Parametric, Implicit, and Explicit Equations

There are basically three types of equations that can be used to define a planar curve: parametric, implicit, and explicit. The parametric equation of a plane curve takes the form

$$x = \frac{x(t)}{w(t)} \qquad y = \frac{y(t)}{w(t)}. \tag{1.28}$$

The implicit equation of a curve is of the form

$$f(x,y) = 0. \tag{1.29}$$

An explicit equation of a curve is a special case of both the parametric and implicit forms:

$$y = f(x). \tag{1.30}$$

In these notes, we restrict ourselves to the case where the functions $x(t)$, $y(t)$, $w(t)$, $f(x)$ and $f(x,y)$ are polynomials.

Any curve that can be expressed parametrically as in equation (1.28) is referred to as a **rational** curve. In the classical algebraic geometry literature, a rational curve is sometimes called a **unicursal** curve, which means that it can be sketched in its entirety without removing one's pencil from the paper. In computer aided geometric design, rational curves are often called **rational parametric** curves. The case where $w(t) \equiv 1$ is called a **polynomial** parametric curve (or a **non-rational** parametric curve). A curve that can be expressed in the form of equation (1.29) is known as a **planar algebraic** curve.

The parametric equation of a curve has the advantage of being able to quickly compute the (x, y) coordinates of points on the curve for plotting purposes. Also, it is simple to define a curve *segment* by restricting the parameter t to a finite range, for example $0 \le t \le 1$. On the other hand, the implicit equation of a curve enables one to easily determine whether a given point lies on the curve, or if not, which side of the curve it lies on. Chapter 17 shows that it is always possible to compute an implicit equation for a parametric curve. It is trivial to convert en explicit equation of a curve into a parametric equation ($x = t$, $y = y(x)$) or into an implicit equation ($f(x) - y = 0$). However, a curve defined by an implicit or parametric equation cannot in general be converted into explicit form.

A **rational surface** is one that can be expressed

$$x = \frac{x(s,t)}{w(s,t)} \qquad y = \frac{y(s,t)}{w(s,t)} \qquad z = \frac{z(s,t)}{w(s,t)} \tag{1.31}$$

where $x(s,t)$, $y(s,t)$, $z(s,t)$ and $w(s,t)$ are polynomials. Also, a surface that can be expressed by the equation

$$f(x,y,z) = 0 \tag{1.32}$$

where $f(x,y,z)$ is a polynomial is called an **algebraic surface.**

A **rational space curve** is one that can be expressed by the parametric equations

$$x = \frac{x(t)}{w(t)} \quad y = \frac{y(t)}{w(t)} \quad z = \frac{z(t)}{w(t)}. \tag{1.33}$$

The curve of intersection of two algebraic surfaces is an **algebraic space curve.**

1.5 Lines

The simplest case of a curve is a line. Even so, there are several different equations that can be used to represent lines.

1.5.1 Parametric equations of lines

Linear parametric equation

A line can be written in parametric form as follows:

$$x = a_0 + a_1 t; \qquad y = b_0 + b_1 t$$

In vector form,

$$\mathbf{P}(t) = \left\{ \begin{array}{c} x(t) \\ y(t) \end{array} \right\} = \left\{ \begin{array}{c} a_0 + a_1 t \\ b_0 + b_1 t \end{array} \right\} = \mathbf{A}_0 + \mathbf{A}_1 t. \tag{1.34}$$

In this equation, $\mathbf{A}_0$ is a point on the line and $\mathbf{A}_1$ is the direction of the line (see Figure 1.7)

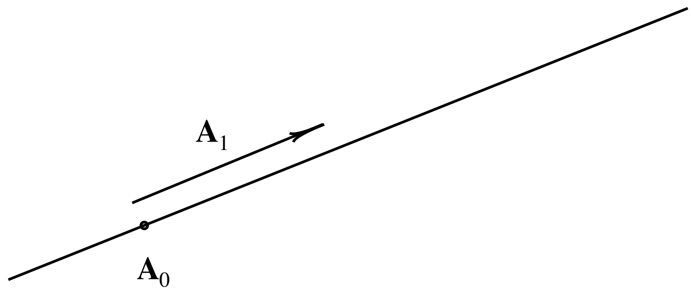
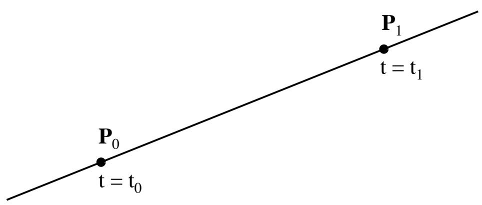
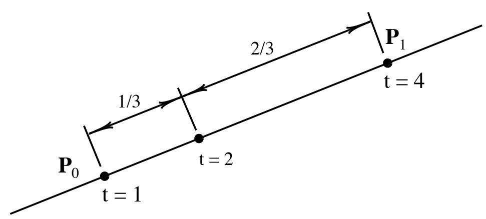


Figure 1.7: Line given by $\mathbf{A}_0 + \mathbf{A}_1 t$.



(a) Line given by $\mathbf{P}(t) = \frac{(t_1-t)\mathbf{P}_0+(t-t_0)\mathbf{P}_1}{t_1-t_0}$.



(b) Affine Example.

Figure 1.8: Affine parametric equation of a line.

Affine parametric equation of a line

A line can also be expressed

$$\mathbf{P}(t) = \frac{(t_1 - t)\mathbf{P}_0 + (t - t_0)\mathbf{P}_1}{t_1 - t_0} \tag{1.35}$$

where $\mathbf{P}_0$ and $\mathbf{P}_1$ are two points on the line and t_0 and t_1 are any parameter values. Note that $\mathbf{P}(t_0) = \mathbf{P}_0$ and $\mathbf{P}(t_1) = \mathbf{P}_1$. Note in Figure 1.8.a that the line *segment* $\mathbf{P}_0$–$\mathbf{P}_1$ is defined by restricting the parameter:

$$t_0 \le t \le t_1.$$

Sometimes this is expressed by saying that the line segment is the portion of the line in the *parameter interval* or *domain* $[t_0, t_1]$.

We will see that the line in Figure 1.8.a is actually a degree one Bézier curve. Most commonly, we have $t_0 = 0$ and $t_1 = 1$ in which case

$$\mathbf{P}(t) = (1 - t)\mathbf{P}_0 + t\mathbf{P}_1. \tag{1.36}$$

Equation 1.36 is called an *affine* equation, whereas equation 1.34 is called a *linear* equation. An affine equation is coordinate system independent, and is mainly concerned with ratios and proportions. An affine equation can be thought of as answering the question: “If a line is defined through two points $\mathbf{P}_0$ and $\mathbf{P}_1$, and if point $\mathbf{P}_0$ corresponds to parameter value t_0 and point $\mathbf{P}_1$ corresponds to parameter value t_1, what point corresponds to an arbitrary parameter value t?” Figure 1.8.b shows a line on which $\mathbf{P}_0$ corresponds to parameter $t = t_0 = 1$ and $\mathbf{P}_1$ is assigned parameter value $t = t_1 = 4$. For example, the point corresponding to $t = 2$ is one third of the way from $\mathbf{P}_0$ to $\mathbf{P}_1$.

Note that an affine equation can be derived from any two points on a line, given the parameter values for those points. If $\mathbf{P}(\alpha)$ is the point corresponding to parameter value $t = \alpha$ and if $\mathbf{P}(\beta)$ is the point corresponding to parameter value $t = \beta$ ($\alpha \neq \beta$), then the point corresponding to parameter value γ is

$$\mathbf{P}(\gamma) = \mathbf{P}(\alpha) + \frac{\gamma - \alpha}{\beta - \alpha}[\mathbf{P}(\beta) - \mathbf{P}(\alpha)] = \frac{(\beta - \gamma)\mathbf{P}(\alpha) + (\gamma - \alpha)\mathbf{P}(\beta)}{\beta - \alpha}$$

Rational parametric equations

A line can also be defined using the following parametric equations:

$$x = \frac{a_0 + a_1 t}{d_0 + d_1 t}; \quad y = \frac{b_0 + b_1 t}{d_0 + d_1 t}. \tag{1.37}$$

This is often called *rational* or *fractional* parametric equations.

Recall that the homogeneous Cartesian coordinates (X, Y, W) of a point are related to its Cartesian coordinates by

$$(x, y) = (\frac{X}{W}, \frac{Y}{W}).$$

Thus, we can rewrite equation 1.37 as

$$X = a_0 + a_1 t; \quad Y = b_0 + b_1 t; \quad W = d_0 + d_1 t.$$

This equation can be further re-written in terms of homogeneous parameters (T, U) where $t = \frac{T}{U}$. Thus,

$$X = a_0 + a_1 \frac{T}{U}; \quad Y = b_0 + b_1 \frac{T}{U}; \quad W = d_0 + d_1 \frac{T}{U}.$$

But since we can scale (X, Y, W) without changing the point (x, y) which it denotes, we can scale by U to give

$$X = a_0 U + a_1 T; \quad Y = b_0 U + b_1 T; \quad W = d_0 U + d_1 T.$$

1.5.2 Implicit equations of lines

A line can also be expressed using an *implicit* equation:

$$f(x, y) = ax + by + c = 0; \quad \text{or} \quad F(X, Y, W) = aX + bY + cW = 0.$$

The line defined by an implicit equation is the set of all points which satisfy the equation $f(x, y) = 0$.

An implicit equation for a line can be derived given a point $\mathbf{P}_0 = (x_0, y_0)$ on the line and the normal vector $\mathbf{n} = a\mathbf{i} + b\mathbf{j}$. As shown in Figure 1.9, a point $\mathbf{P} = (x, y)$ is on this line if

$$(\mathbf{P} - \mathbf{P}_0) \cdot \mathbf{n} = 0$$

from which

$$f(x, y) = (x - x_0, y - y_0) \cdot (a, b) = ax + by - (ax_0 + by_0) = 0. \tag{1.38}$$

From equation 1.38, a line whose implicit equation is $ax + by + c = 0$ has the normal vector $\mathbf{n} = a\mathbf{i} + b\mathbf{j}$.

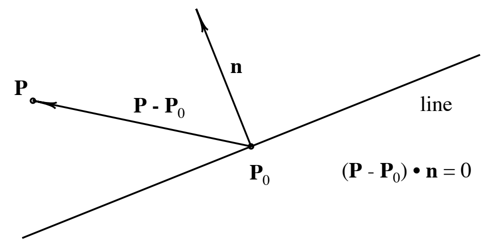


Figure 1.9: Line defined by point and normal.

Implicit equation of a line through two points

Three points (X_1, Y_1, W_1), (X_2, Y_2, W_2) and (X_3, Y_3, W_3) are collinear if

$$\begin{vmatrix} X_1 & Y_1 & W_1 \\ X_2 & Y_2 & W_2 \\ X_3 & Y_3 & W_3 \end{vmatrix} = 0.$$

Thus, the equation of the line through two points is

$$\begin{vmatrix} X & Y & W \\ X_1 & Y_1 & W_1 \\ X_2 & Y_2 & W_2 \end{vmatrix} = (Y_1W_2 - Y_2W_1)X + (X_2W_1 - X_1W_2)Y + (X_1Y_2 - X_2Y_1)W = 0.$$

1.5.3 Distance from a point to a line

If $\mathbf{n} = a\mathbf{i} + b\mathbf{j}$ is a unit vector (that is, if $a^2 + b^2 = 1$), then the value $f(x, y)$ in equation 1.38 indicates the signed perpendicular distance of a point (x, y) to the line. This can be seen from equation 1.38 and Figure 1.9. The dot product $(\mathbf{P} - \mathbf{P}_0) \cdot \mathbf{n}$ is the length of the projection of vector $(\mathbf{P} - \mathbf{P}_0)$ onto the unit normal $\mathbf{n}$, which is the perpendicular distance from $\mathbf{P}$ to the line.

Since the coefficients of an implicit equation can be uniformly scaled without changing the curve (because if $f(x, y) = 0$, then $c \times f(x, y) = 0$ also), the implicit equation of a line can always be *normalized*:

$$f(x, y) = a'x + b'y + c' = \frac{a}{\sqrt{a^2 + b^2}}x + \frac{b}{\sqrt{a^2 + b^2}}y + \frac{c}{\sqrt{a^2 + b^2}} = 0.$$

Then, $f(x, y)$ is the signed distance from the point (x, y) to the line, with all points on one side of the line having $f(x, y) > 0$ and the other side having $f(x, y) < 0$. Note that $|c'| = |f(0, 0)|$ is the distance from the origin to the line. Thus, if $c = 0$, the line passes through the origin. The coefficients a' and b' relate to the slope of the line. Referring to Figure 1.10, $a' = \cos(\theta)$, $b' = \sin(\theta)$, and $c' = -p$.

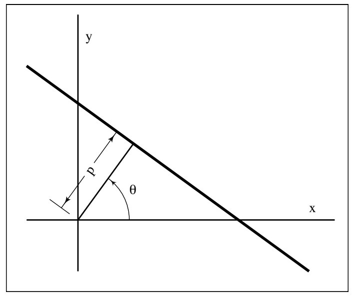


Figure 1.10: Normalized line equation.

1.6 Conic Sections

A conic section (or, simply *conic*) is any degree two curve. Any conic can be expressed using a degree two implicit equation:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

or, in homogeneous form:

$$aX^2 + bXY + cY^2 + dXW + eYW + fW^2 = 0. \tag{1.39}$$

Conics can be classified as hyperbolas, parabolas and ellipses (of which the circle is a special case). What distinguishes these cases is the number of real points at which the curve intersects the line at infinity $W = 0$. A hyperbola intersects $W = 0$ in two real points. Those points are located an infinite distance along the asymptotic directions. A parabola is tangent to the line at infinity, and thus has two coincident real intersection points. This point is located an infinite distance along the parabola's axis of symmetry. Ellipses do not intersect the line at infinity at any real point — all real points on an ellipse are finite.

To determine the number of real points at which a conic intersects the line at infinity, simply intersect equation 1.39 with the line $W = 0$ by setting $W = 0$ to get:

$$aX^2 + bXY + cY^2 = 0$$

from which

$$\frac{Y}{X} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c}.$$

The two values Y/X are the slopes of the lines pointing to the intersections of the conic with the line at infinity. Thus, if $b^2 - 4ac > 0$, there are two distinct real intersections and the conic is a hyperbola. If $b^2 - 4ac = 0$, there are two coincident real intersections and the conic is a parabola, and if $b^2 - 4ac < 0$, there are no real intersections and the conic is an ellipse. The value $b^2 - 4ac$ is known as the *discriminant* of the conic.

1.6.1 Parametric equations of conics

The parametric equation of any conic can be expressed:

$$x = \frac{a_2t^2 + a_1t + a_0}{d_2t^2 + d_1t + d_0}; \qquad y = \frac{b_2t^2 + b_1t + b_0}{d_2t^2 + d_1t + d_0}.$$

or, in homogeneous form,

$$X = a_2T^2 + a_1TU + a_0U^2;$$
$$Y = b_2T^2 + b_1TU + b_0U^2;$$
$$W = d_2T^2 + d_1TU + d_0U^2.$$

It is also possible to classify a conic from its parametric equation. We again identify the points at which the conic intersects the line at infinity. In the parametric form, the only places at which (x, y) can be infinitely large is at parameter values of t for which

$$d_2t^2 + d_1t + d_0 = 0.$$

Thus, we note that $d_1^2 - 4d_0d_2$ serves the same function as the discriminant of the implicit equation. If $d_1^2 - 4d_0d_2 > 0$, there are two real, distinct values of t at which the conic goes to infinity and the curve is a hyperbola. If $d_1^2 - 4d_0d_2 < 0$, there are no real values of t at which the conic goes to infinity and the curve is an ellipse. If $d_1^2 - 4d_0d_2 = 0$, there are two real, identical values of t at which the conic goes to infinity and the curve is a parabola.