Mech 289 Design Graphics

Module 2: Fundamentals of Geometry Construction

Department of Mechanical Engineering
McGill University, Montreal
Canada

December 2006

Contents

List of Figures

Chapter 1

Introduction to Geometry Construction

Beware of designers who walk around
without paper and pencil.
Ancient Chinese proverb.

1.1 Overview of Module 2

The design process starts with a need, as spelled out by the client. In engineering design as well as in other design areas, the need is described by the client in rather ambiguous, fuzzy, sometimes contradictory terms. After a series of exchanges between client and designer, be this an engineer, an industrial designer or an architect, the need is formulated in terms of a list of *functional requirements*, with some specific features that are spelled out as *design specifications*, or *specs* for brevity.

Once the functional requirements and design specifications are agreed upon by client and designer, the latter produces free-hand sketches of some design alternatives. The importance of free-hand sketching skills in design cannot be overstated. The quality of the final design solution is highly dependent upon how the designer can communicate her or his ideas not only to the client and to other professionals, but also to herself or himself in an unambiguous, concise and clear way. Developing basic sketching skills is the subject of Module 1.

Shown in Fig. 1.1 is a free-hand sketch produced by a mechanical designer *to embody* the design of a mechanism housing, to serve as a means to protect the mechanism, displayed in Fig. 1.2 as mounted on a jig, and to support firmly its various moving parts.

Module 2 aims at developing the analytical skills required in producing accurate, unambiguous engineering drawings, to be used by other engineers and technicians for manufacturing or construction. This module relies on elementary knowledge of algebra and geometry as prerequisites. The main objective of the course is to teach students the "math behind the CAD," where CAD stands for *Computer-Aided Design*, a technology that frees the designer from the routine tasks of design drafting and modelling. Students in this course will be introduced to CAD in Module 3.

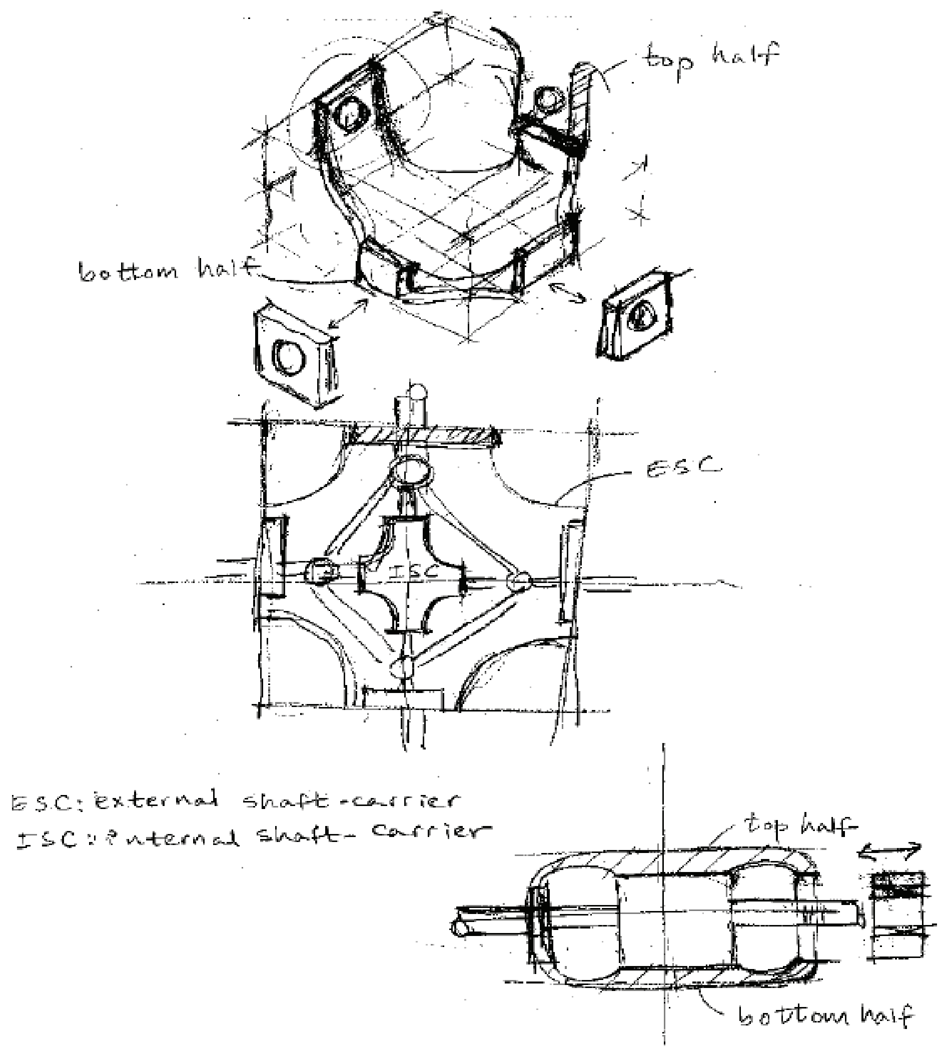


Figure 1.1: The free-hand sketch of a solution alternative to the problem of designing a mechanism housing

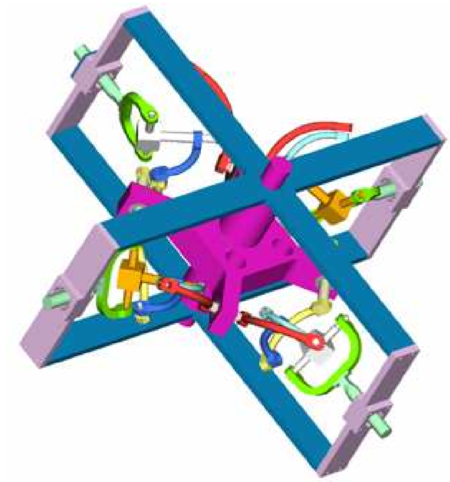


Figure 1.2: The geometric modelling of a complex mechanism, mounted on a jig, in need of a housing

A manufacturing or construction drawing is, as a rule, a 2D drawing displaying the relevant dimensions and other information—e.g., tolerances and materials—required to produce parts or full design solutions. Sometimes, an isometric view of the part in question is added for visualization purposes. Illustrated in Fig. 1.3 is one manufacturing drawing that shows one part of the mechanism housing sketched in Fig. 1.1.

A geometric model of the design solution is a more realistic representation of the same object, intended not only for visualization, but also for the calculation of the various geometric properties—volume, footprint area, centroid location, moment of inertia—and mechanical behaviour of the object. The latter can include stress and vibration analyses by means of CAE, an acronym for *Computer-Aided Engineering*. A geometric model of the same housing, with all its parts assembled, although not including the mechanism it is intended to hold, is illustrated in Fig. 1.4.

1.2 Back to Basics: Coordinate Space

1.2.1 The Cartesian Coordinate System

In order to locate points, lines, planes, or other geometric objects in space, the positions—and/or their orientation, as the case may be—of these objects must be known with respect to some reference frame. Generally, we use the *Cartesian* coordinate system to allow the position and orientation of geometric objects to be referenced relative to a selected frame.

As illustrated in Fig. 1.5, a 2-dimensional coordinate system establishes an origin at the intersection of two mutually perpendicular axes, conventionally labeled X (horizontal) and Y (vertical). The origin is assigned the coordinate values of $(0, 0)$.

Using this coordinate system, we are able to construct a multitude of geometric objects by specifying the coordinates of the vertices and connecting them together with lines to form edges. An example of this is the rectangle shown in Fig. 1.6.

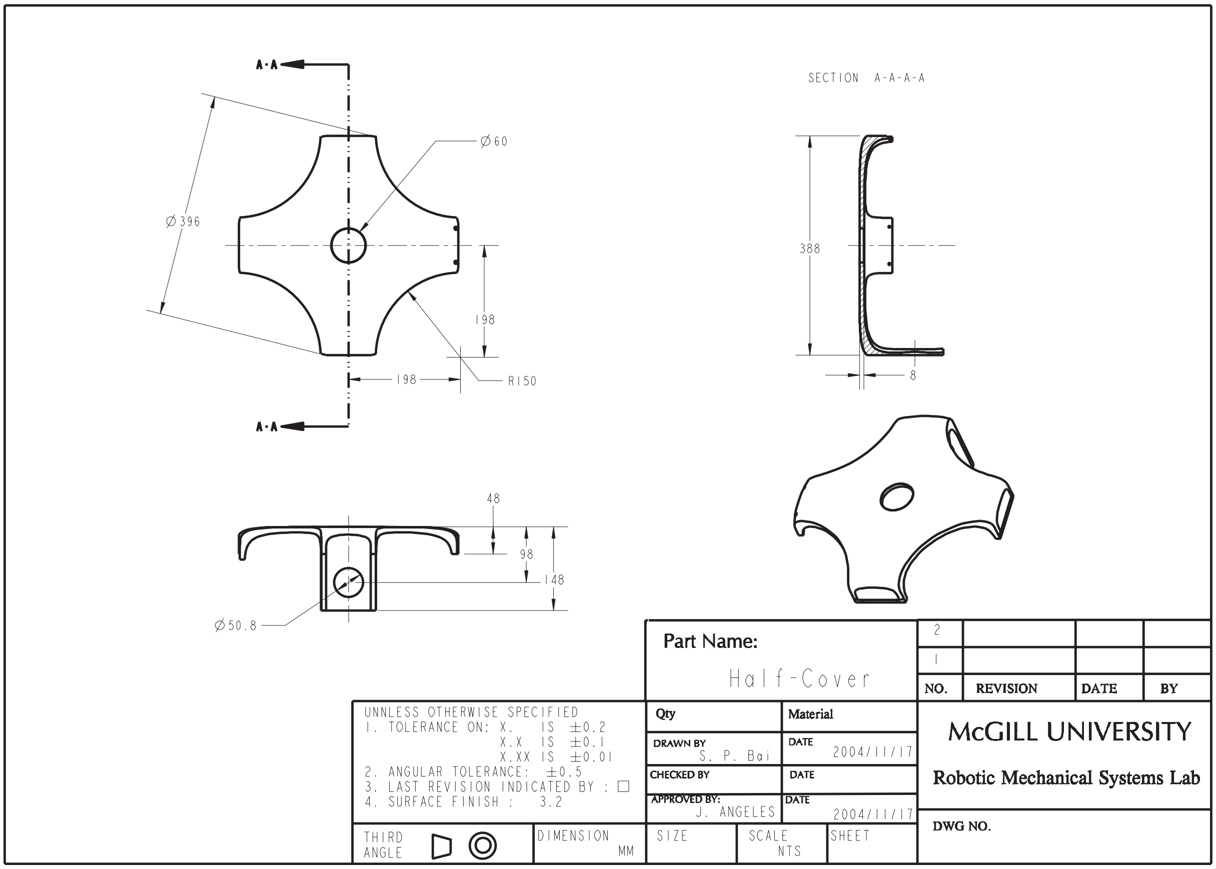


Figure 1.3: A typical manufacturing drawing of a part of the housing sketched in Fig. 1.1, with a (non-isometric) 3D view added for visualization

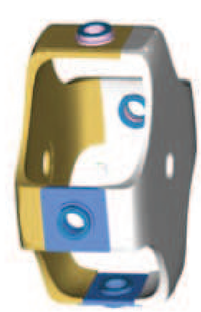


Figure 1.4: A geometric model of the mechanism housing of Fig. 1.3

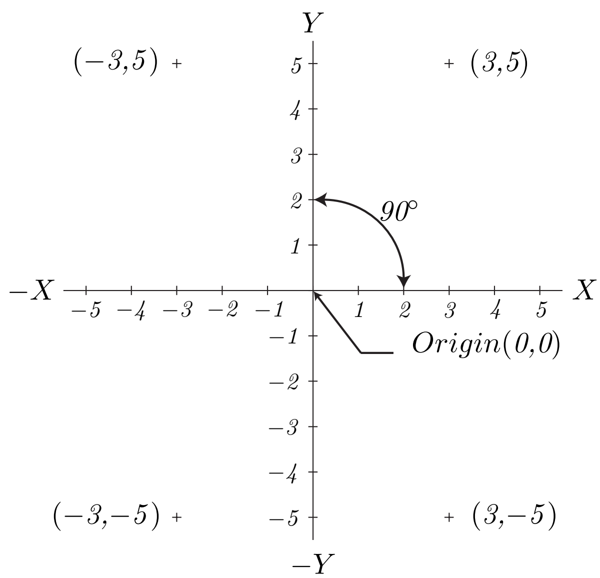


Figure 1.5: A 2D coordinate system

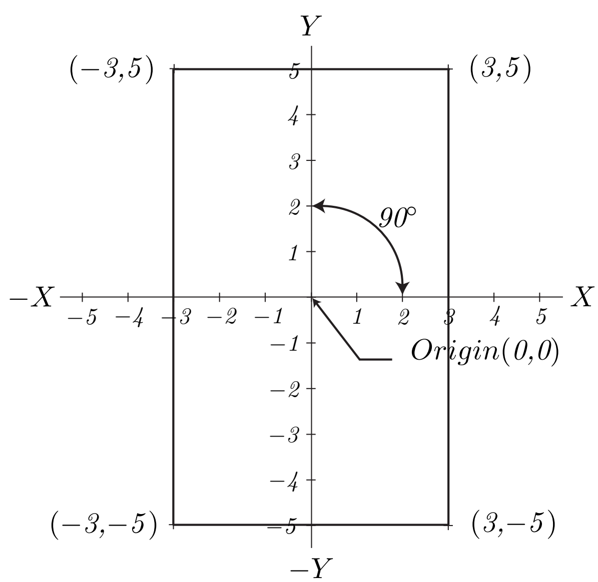


Figure 1.6: Creating a rectangle

However, even though a 2D system may be useful for a variety of purposes, most real-world applications require a third dimension. The 2D Cartesian plane can readily be extended to allow for the inclusion of 3D points. In a 3D coordinate system, the origin is established at the point where three mutually perpendicular axes—X, Y and Z—intersect. The origin is assigned the coordinate values of $(0, 0, 0)$, as illustrated in Fig. 1.7.

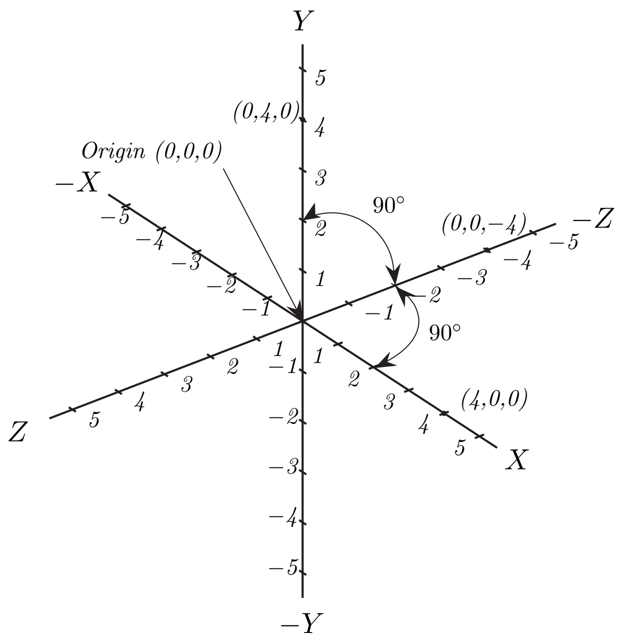


Figure 1.7: 3D coordinate system

Similar to the 2D rectangle that was constructed, a rectangular prism is created using the 3D coordinate system by establishing coordinate values for each vertex, and then inserting the appropriate edges, as shown in Fig. 1.8.

This coordinate system is used in multiview drawings and 3D modelling, using both traditional tools and Computer Aided Design (CAD) tools. Figure 1.9 is a multiview drawing of an object, with coordinate axes displayed in each viewport. Only two of the three coordinates can be seen in each view.

CAD systems provide a method for displaying the current position of the cursor in a coordinate frame, as shown on the screen snapshot in Fig. 1.10.

1.2.2 Right-Hand Rule

The right-hand rule is used to determine the positive direction of the axes; it defines the X, Y and Z axes as well as the positive and negative directions of rotation on each axis.

As illustrated in Fig. 1.11, the simplest way to remember the right-hand rule is to first make a fist with your right hand, ensuring that your thumb points outward. The direction in which your thumb points is the positive direction of the X-axis. Straighten your index finger so that it points straight up, at a 90° degree angle to your thumb. The direction of your

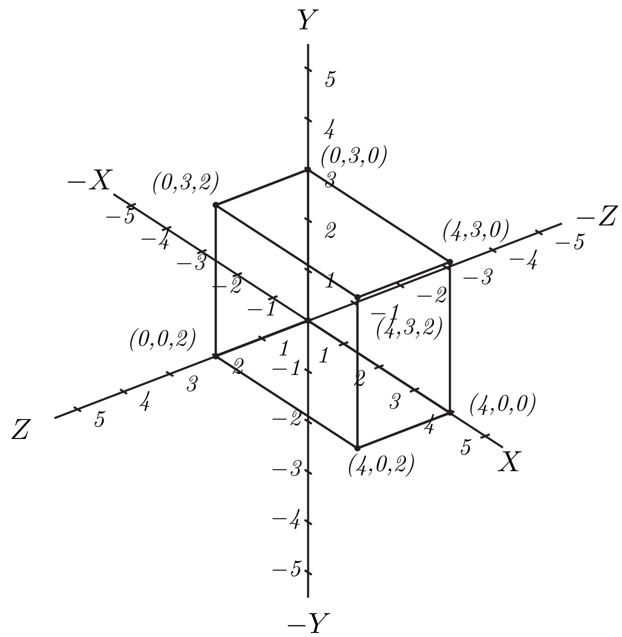


Figure 1.8: Creating a rectangular parallelepiped

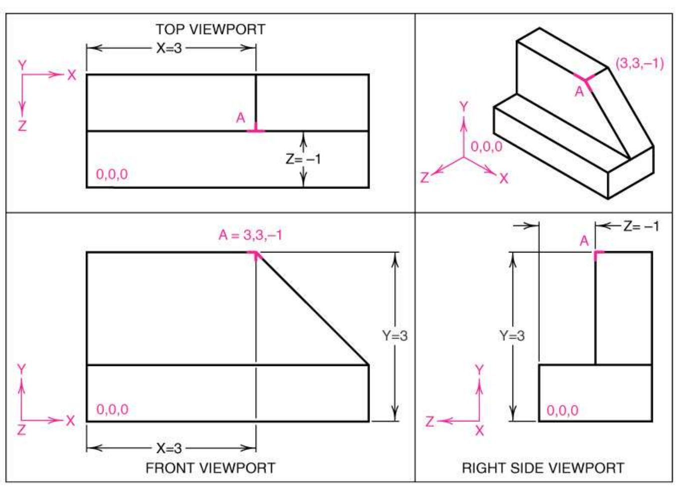


Figure 1.9: Display of coordinate axes in a multiview CAD drawing

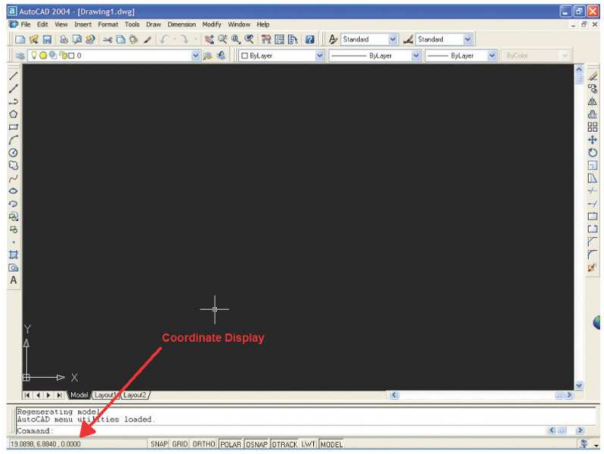


Figure 1.10: Cursor on a CAD screen

index finger indicates the positive direction on the Y-axis. Similarly, straighten your middle finger so that it points straight up, at 90° to your index finger. The direction of your middle finger indicates the positive direction on the Z-axis.

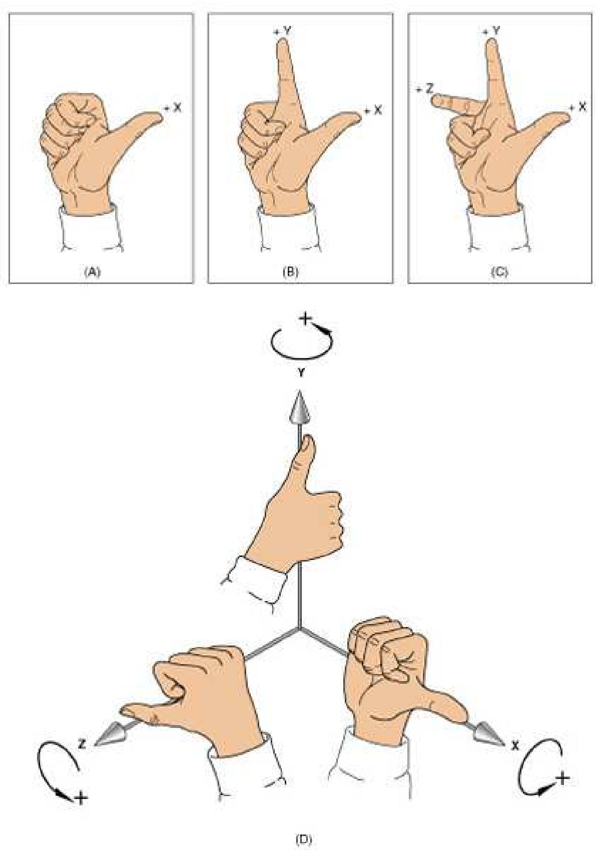


Figure 1.11: Right Hand Rule

It is noteworthy that opposite to the right-hand rule, there exists a left-hand rule which defines all of the previous axes in the same way as the right-hand rule, only *reflected*. While the left-hand rule is used in some situations to describe the coordinate axes, in this course we will only employ the conventional right-hand rule for the sake of consistency, simplicity and established convention.

1.2.3 Types of Coordinate Systems

Polar Coordinates

Polar coordinates are used to locate points in the plane; they specify a distance and an angle from the origin $(0, 0)$. Figure 1.12 shows a line in the XY-plane, 4.5 units long and at an angle of 30° from the X-axis. Polar coordinates are commonly used by CAD systems to locate points because of their inherent simplicity.

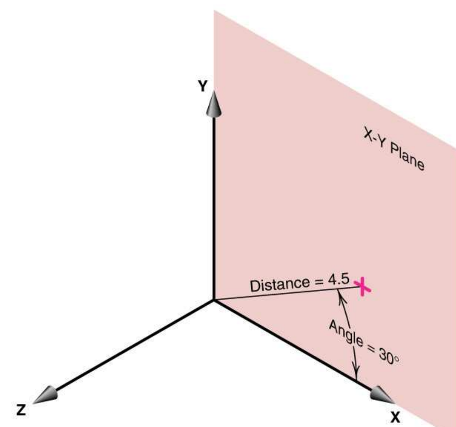


Figure 1.12: Polar coordinates

Cylindrical Coordinates

Cylindrical coordinates locate a point on the surface of a cylinder by specifying a distance from the origin and an angle from the X-axis in the XY-plane, and the distance in the direction of Z. In Fig. 1.13, for example, point A is a distance z from the XY plane, a distance r from the origin as measured on a line that makes an angle θ with the X-axis and lying in the XY-plane. Notice that the radius of the cylinder corresponds to the radius of the cylindrical coordinate in question, with the point located on the surface of the cylinder itself.

In general, cylindrical coordinates are used in designing axially symmetric shapes.

To change cylindrical coordinates to Cartesian coordinates, use the following transformation:

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= z \end{aligned} \tag{1.1}$$

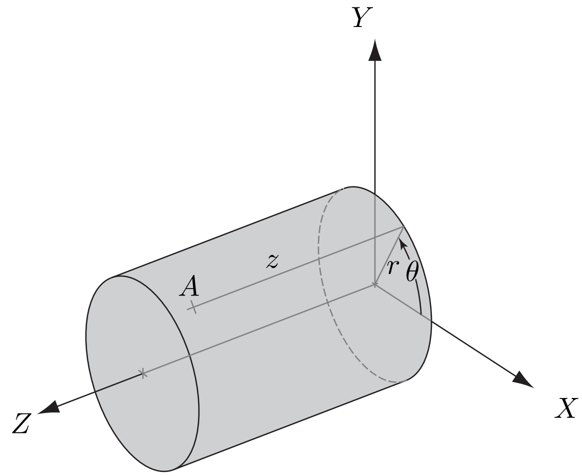


Figure 1.13: Cylindrical coordinates

Spherical Coordinates

Spherical coordinates locate a point on the surface of a sphere by specifying an angle θ in the XY-plane, an angle ϕ in plane making a dihedral angle θ with the XZ-plane, and one radial distance r, as shown in Fig. 1.14.

To change spherical coordinates to Cartesian coordinates, we use the transformation below:

$$
\begin{aligned}
x &= r\cos\phi\cos\theta \\
y &= r\cos\phi\sin\theta \\
z &= r\sin\phi
\end{aligned}
\tag{1.2}
$$

As an exercise, derive the equations to transform (x, y, z) coordinates into their spherical counterparts, (r, θ, ϕ). **Hint:** Think of Pythagoras and your trigonometry skills.

Absolute/Relative/World Coordinate Systems

As illustrated in Fig. 1.15, *absolute coordinates* always refer to the origin $(0, 0, 0)$.

*Relative coordinate*s are always referenced to a previously defined location and are sometimes referred to as Δ-coordinates, as shown in Fig. 1.16. Thus, you can have several different coordinate systems within one larger coordinate system; each object may have its own local (relative) system, but each local system is referenced to an encompassing “world” coordinate system.

As illustrated in Fig. 1.17, the *world coordinate system* uses a set of three numbers (x, y, z) located on three mutually perpendicular axes and measured from the origin $(0, 0, 0)$; the local coordinate system is a moving system that can be positioned anywhere in 3D space to assist in the construction of geometric objects.

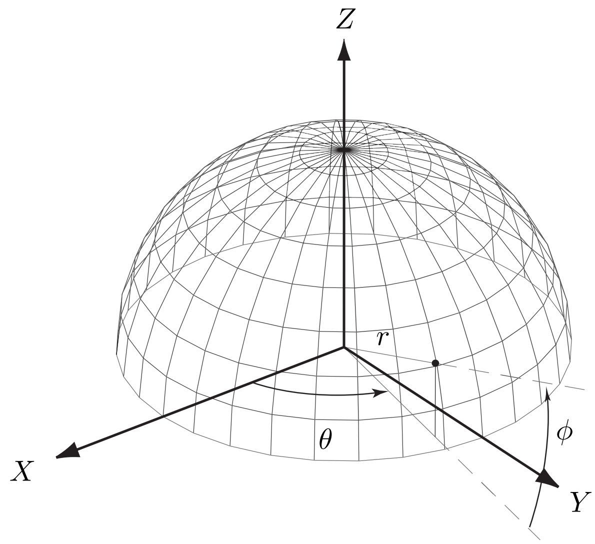


Figure 1.14: Spherical coordinates

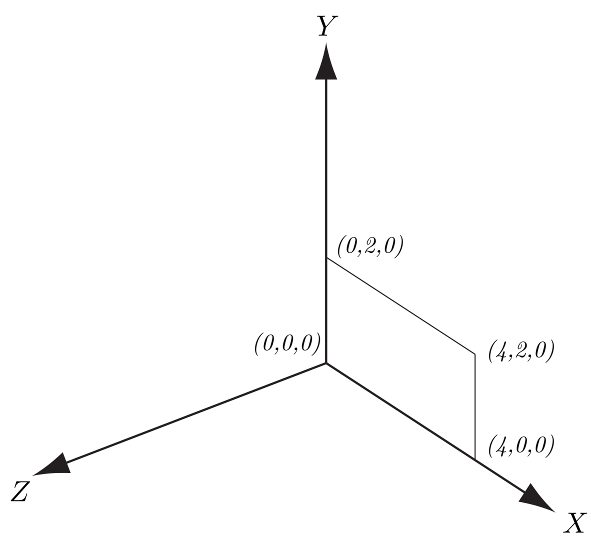


Figure 1.15: Absolute coordinates

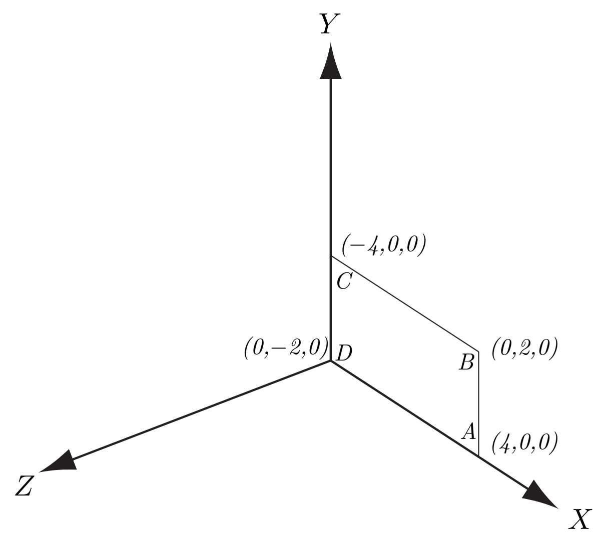


Figure 1.16: Relative coordinates

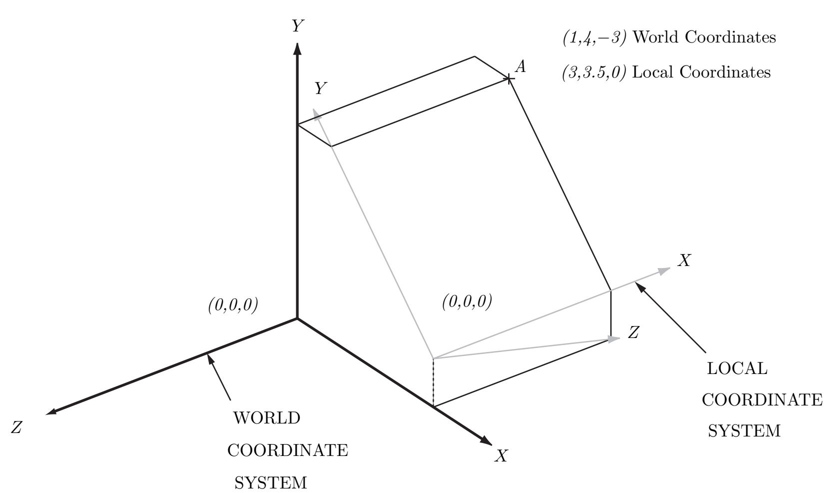


Figure 1.17: World and Local coordinates

1.2.4 Homogeneous Coordinates

One of the many purposes of using *homogeneous coordinates* is to capture the concept of infinity. In the Euclidean coordinate system, infinity is something that does not exist. Mathematicians have discovered that many geometric concepts and computations can be greatly simplified if the concept of infinity is used. This will become apparent when we move to curves and surface design. Without the use of a homogeneous coordinates system, it would be difficult to design certain classes of frequently used curves and surfaces in computer graphics and computer-aided design, as well as perform transformations on these curves and surfaces.

Any point in the XY-plane has two coordinates. Adding a third component whose value is 1 to the coordinates of this point leads to the corresponding homogenous coordinates. Thus, the homogeneous coordinates of any point P in the said plane are:

$$\boldsymbol{p} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1.3}$$

A point P_∞ lying at infinity, with a line of sight making an angle θ with the X-axis, has the homogeneous coordinates stored in the array $\boldsymbol{p}_\infty$ given below:

$$\boldsymbol{p}_\infty = \begin{bmatrix} \cos\theta \\ \sin\theta \\ 0 \end{bmatrix} \tag{1.4}$$

Similarly, the homogeneous coordinates of a point in 3D are defined as:

$$\boldsymbol{p} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{1.5}$$

Likewise, the homogeneous coordinates of a point P_∞ lying at infinity in 3D space, with a line of sight of direction cosines (λ, μ, ν), are stored in an array $\boldsymbol{p}_\infty$ given as

$$\boldsymbol{p}_\infty = \begin{bmatrix} \lambda \\ \mu \\ \nu \\ 0 \end{bmatrix} \tag{1.6}$$

Homogeneous coordinates have been traditionally used in place of ordinary Cartesian coordinates in computer graphics and geometric modelling. The representation of points in homogeneous coordinates provides a unified approach to the description of geometric transformations, and allows these transformations to be represented as simple matrix operations.

In this course, we will utilize homogeneous coordinates in Chapter 4 when we will study *affine transformations*. As we will see, homogenous coordinates simplify (and in many cases, make possible) the mathematics needed to represent the desired transformations and manipulations.

1.3 Vectors

1.3.1 Notation

Throughout this work, we use boldface fonts to indicate vectors ($\boldsymbol{a}$) and matrices ($\boldsymbol{R}$), with uppercase letters reserved for matrices and lowercases for vectors. Additionally, calligraphic literals ($\mathcal{C}$) are reserved for sets of points or of other objects.

1.3.2 Definition

A vector is a mathematical entity which has:

- A direction
- An orientation
- A norm (or magnitude)

A vector has a tail and a head, which determines its orientation. Conventionally, the head of a vector is indicated by an arrow, and the tail by a point. These two are joined together by a line to form the vector.

We now introduce the *unit vectors* $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$: vector $\boldsymbol{i}$ is parallel to the X-axis, $\boldsymbol{j}$ parallel to the Y-axis,and $\boldsymbol{k}$ parallel to the Z-axis. These vectors are given in component form as

$$\boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \boldsymbol{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{1.7}$$

Because we can multiply a vector by some scalar quantity that changes its magnitude but not its direction, we can express any given vector $\boldsymbol{v}$ as:

$$\boldsymbol{v} = v_x\boldsymbol{i} + v_y\boldsymbol{j} + v_z\boldsymbol{k} \tag{1.8}$$

This follows from $\boldsymbol{v} = \boldsymbol{v}_x + \boldsymbol{v}_y + \boldsymbol{v}_z$, where $\boldsymbol{v}_x = v_x\boldsymbol{i}$, $\boldsymbol{v}_y = v_y\boldsymbol{j}$, and $\boldsymbol{v}_z = v_z\boldsymbol{k}$.

The same vector $\boldsymbol{v}$ can be described as a *column array*:

$$\boldsymbol{v} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} \tag{1.9}$$

where v_x, v_y, v_z are the components of $\boldsymbol{v}$. The components may be negative, depending on the direction of the vector.

1.3.3 Basic Properties

Magnitude of a vector

The magnitude of a vector, also known as the *Euclidean vector norm*, is non-negative, and vanishes only when the vector itself does. The magnitude of $\boldsymbol{v}$ is thus a non-negative scalar quantity, denoted by $\|\boldsymbol{v}\|$ and given by:

$$\|\boldsymbol{v}\| = \sqrt{v_x^2 + v_y^2 + v_z^2} \tag{1.10}$$

which is a simple application of the Pythagorean theorem used to find the length of the diagonal of a parallelepiped of sides with lengths v_x, v_y, v_z, such as the one shown in Fig. 1.8. Hence,

$$\|\boldsymbol{v}\|^2 = v_x^2 + v_y^2 + v_z^2 \tag{1.11}$$

Unit vector

We define a unit vector as any vector whose magnitude is equal to unity, regardless of its direction. As we saw, $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ are special cases of unit vectors, with specific directions assigned to them. An arbitrary unit vector in the direction of $\boldsymbol{v}$ can be obtained as:

$$\boldsymbol{w} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \tag{1.12}$$

with

$$\|\boldsymbol{w}\| = 1 \tag{1.13}$$

A unit vector can be also written in the form

$$\boldsymbol{w} = \begin{bmatrix} v_x/\|\boldsymbol{v}\| \\ v_y/\|\boldsymbol{v}\| \\ v_z/|\boldsymbol{v}\| \end{bmatrix} \tag{1.14}$$

We can make this more concise with the definitions:

$$w_x = \frac{v_x}{\|\boldsymbol{v}\|},\ w_y = \frac{v_y}{\|\boldsymbol{v}\|},\ w_z = \frac{v_z}{\|\boldsymbol{v}\|} \tag{1.15}$$

so that:

$$\boldsymbol{w} = \begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix} \tag{1.16}$$

Note that if α, β, and γ are the angles between $\boldsymbol{v}$ and the X, Y and Z-axes, respectively, then:

$$w_x = \frac{v_x}{\|\boldsymbol{v}\|} = \cos\alpha \qquad w_y = \frac{v_y}{\|\boldsymbol{v}\|} = \cos\beta \qquad w_z = \frac{v_z}{\|\boldsymbol{v}\|} = \cos\gamma \tag{1.17}$$

which indicates that w_x, w_y, w_z are the *direction cosines* of $\boldsymbol{v}$.

Scalar multiplication

Multiplying any vector $\boldsymbol{v}$ by a scalar k produces a vector $k\boldsymbol{v}$:

$$k\boldsymbol{v} = \begin{bmatrix} kv_x \\ kv_y \\ kv_z \end{bmatrix} \tag{1.18}$$

If k is positive, then $\boldsymbol{v}$ and $k\boldsymbol{v}$ are in the same direction; if k is negative, then $\boldsymbol{v}$ and $k\boldsymbol{v}$ are in opposite directions. The magnitude of $k\boldsymbol{v}$ is:

$$\|k\boldsymbol{v}\| = \sqrt{k^2v_x^2 + k^2v_y^2 + k^2v_z^2} \tag{1.19}$$

so that

$$\|k\boldsymbol{v}\| = |k|\|\boldsymbol{v}\| \tag{1.20}$$

Vector addition

Given $\boldsymbol{a} = [a_x\ a_y\ a_z]^T$ and $\boldsymbol{b} = [b_x\ b_y\ b_z]^T$, the sum of these two vectors is defined as

$$\boldsymbol{a} + \boldsymbol{b} = \begin{bmatrix} a_x + b_x \\ a_y + b_y \\ a_z + b_z \end{bmatrix} \tag{1.21}$$

Given vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ and scalars k and l, vector addition and scalar multiplication obey the properties below:

1. $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$
2. $\boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c}) = (\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c}$
3. $k(l\boldsymbol{a}) = kl\boldsymbol{a}$
4. $(k + l)\boldsymbol{a} = k\boldsymbol{a} + l\boldsymbol{a}$
5. $k(\boldsymbol{a} + \boldsymbol{b}) = k\boldsymbol{a} + k\boldsymbol{b}$

1.3.4 Scalar Product

The scalar product, also known as *dot product*, of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is the sum of the products of their corresponding components:

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_xb_x + a_yb_y + a_zb_z \tag{1.22}$$

which returns a scalar quantity, and not another vector. An alternative form representing the scalar product is

$$\boldsymbol{a}^T\boldsymbol{b} = a_xb_x + a_yb_y + a_zb_z \tag{1.23}$$

In addition, we can readily show that the scalar product is commutative, meaning that:

$$\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a} \tag{1.24}$$

Alternatively, the scalar product may be calculated using the angle θ between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$\boldsymbol{a} \cdot \boldsymbol{b} = \|\boldsymbol{a}\| \|\boldsymbol{b}\| \cos\theta \tag{1.25}$$

Moreover, the following two statements are equivalent:

$$\boldsymbol{a} \cdot \boldsymbol{b} = 0 \Longleftrightarrow \boldsymbol{a} \,\text{and}\, \boldsymbol{b} \,\text{are perpendicular.} \tag{1.26}$$

In summary, the scalar product has the properties below:

1. $\boldsymbol{a} \cdot \boldsymbol{b} = \|\boldsymbol{a}\| \|\boldsymbol{b}\| \cos\theta$, where θ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$
2. $\boldsymbol{a} \cdot \boldsymbol{a} = \|\boldsymbol{a}\|^2$
3. $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a}$, *commutativity*
4. $\boldsymbol{a} \cdot (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{a} \cdot \boldsymbol{c}$, *distributivity*
5. $(k\boldsymbol{a}) \cdot \boldsymbol{b} = \boldsymbol{a} \cdot (k\boldsymbol{b}) = k(\boldsymbol{a} \cdot \boldsymbol{b})$, *associativity*
6. $\boldsymbol{a}$ is perpendicular to $\boldsymbol{b} \Longleftrightarrow \boldsymbol{a} \cdot \boldsymbol{b} = 0$

1.3.5 Vector Product

The vector product of two 3D vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, also known as the *cross product*, is defined as:

$$\boldsymbol{a} \times \boldsymbol{b} = (a_y b_z - a_z b_y)\boldsymbol{i} - (a_x b_z - a_z b_x)\boldsymbol{j} + (a_x b_y - a_y b_x)\boldsymbol{k} \tag{1.27}$$

In array form,

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{bmatrix} a_y b_z - a_z b_y \\ a_z b_x - a_x b_z \\ a_x b_y - a_y b_x \end{bmatrix} \tag{1.28}$$

One mnemonic means to compute the vector product relies on the expansion of a determinant (cofactor expansion, as outlined in Subsection 1.4.5), namely:

$$\begin{aligned} \boldsymbol{a} \times \boldsymbol{b} &= \det \begin{bmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{bmatrix} \\ &= (a_y b_z - a_z b_y)\boldsymbol{i} - (a_z b_x - a_x b_z)\boldsymbol{j} + (a_x b_y - a_y b_x)\boldsymbol{k} \end{aligned} \tag{1.29}$$

The properties of the vector product follow:

- If $\boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b}$, then $\boldsymbol{c}$ is perpendicular to both $\boldsymbol{a}$ and $\boldsymbol{b}$. Consequently, $\boldsymbol{c}$ is also perpendicular to the plane defined by $\boldsymbol{a}$ and $\boldsymbol{b}$.
- The vector product is *skew-symmetric*: $\boldsymbol{a} \times \boldsymbol{b} = -(\boldsymbol{b} \times \boldsymbol{a})$
- $\boldsymbol{a}$ and $\boldsymbol{b}$ are parallel $\Longleftrightarrow \boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{0}$,

Also note that, if $\boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b}$, and θ denotes the angle from $\boldsymbol{a}$ to $\boldsymbol{b}$, in this direction, when the tails of the two vectors coincide, then

$$\boldsymbol{c} = (\|\boldsymbol{a}\|\|\boldsymbol{b}\| \sin\theta)\boldsymbol{n} \tag{1.30}$$

where $\boldsymbol{n}$ is the unit vector normal to both $\boldsymbol{a}$ and $\boldsymbol{b}$, so that $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{n})$ is a right-hand triad.

1.3.6 Inequalities

In connection with vector norms, there are two important inequalities that arise:

- Cauchy-Schwartz inequality:
$$(\boldsymbol{v} \cdot \boldsymbol{w})^2 \leq \|\boldsymbol{v}\|^2 \|\boldsymbol{w}\|^2 \tag{1.31}$$
- Triangle inequality:
$$\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\| \tag{1.32}$$

These properties are quite useful for the derivation of other identities and inequalities, and are fundamental to the understanding and application of vectors in computer graphics.

1.4 Matrices

1.4.1 Definition

A matrix is a rectangular array of numbers arranged in m rows and n columns, namely,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \tag{1.33}$$

1.4.2 Special Matrices

- Square matrix:

  A square matrix has an equal number of rows and columns ($m = n$), e.g.,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{1.34}$$

- Row matrix:

  A row matrix has one row:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \tag{1.35}$$

- Column matrix:

  A column matrix has one column. In our choice of notation, this column matrix becomes a vector array, namely,

$$\boldsymbol{b} = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} \tag{1.36}$$

  Note the use of a bold *lowercase* letter, indicating that this is a vector.

- Diagonal matrix:

  A square matrix that has zero entries everywhere except on the main diagonal is termed *diagonal*:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & 0 & 0 & 0 & \ldots & 0 \\ 0 & a_{22} & 0 & 0 & \ldots & \vdots \\ 0 & 0 & a_{33} & 0 & \ldots & \vdots \\ 0 & 0 & 0 & a_{44} & \ldots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & \ldots & \ldots & \ldots & 0 & a_{mm} \end{bmatrix} \tag{1.37}$$

  Diagonal matrices satisfy: $a_{ij} = 0$ if $i \neq j$.

- Identity matrix:

  This is a special diagonal matrix with unit elements on the main diagonal, and zero everywhere else. It is denoted by the symbol **1**. For example, the 3×3 identity matrix is defined as

$$\mathbf{1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.38}$$

  The elements of **1** are sometimes denoted by the *Kronecker delta*, namely,

$$\begin{aligned} \delta_{ij} &= 0 \quad \text{if } i \neq j \\ \delta_{ij} &= 1 \quad \text{if } i = j \end{aligned} \tag{1.39}$$

- Zero matrix:

  The $m \times n$ zero matrix has all its elements equal to zero. We will denote it by $\boldsymbol{O}$.

$$\boldsymbol{O} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{1.40}$$

- Symmetric matrix:

  Symmetric matrices are symmetric about the main diagonal: $a_{ij} = a_{ji}$, i.e.

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ \vdots & a_{22} & a_{23} \\ \text{symmetric} & \ldots & a_{33} \end{bmatrix} \tag{1.41}$$

- Skew-symmetric matrix:

  A matrix is skew-symmetric if $a_{ij} = -a_{ji}$, namely,

$$\boldsymbol{A} = \begin{bmatrix} 0 & a_{12} & a_{13} \\ -a_{12} & 0 & a_{23} \\ -a_{13} & -a_{23} & 0 \end{bmatrix} \tag{1.42}$$

- Triangular matrix: An *upper-triangular* matrix has all its entries below the diagonal equal to zero, namely,

$$\boldsymbol{U} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ 0 & a_{21} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{nn} \end{bmatrix} \tag{1.43}$$

  A *lower-triangular* matrix is defined correspondingly.

1.4.3 Properties

- Matrix equality: Let us assume that Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same numbers of rows and columns. Then,

$$\boldsymbol{A} = \boldsymbol{B} \Longleftrightarrow a_{ij} = b_{ij}, \text{ for all } i,j. \tag{1.44}$$

- Matrix addition:

  Adding two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ produces a third matrix $\boldsymbol{C}$, whose elements are equal to the sum of the corresponding elements of $\boldsymbol{A}$ and $\boldsymbol{B}$. Thus,

$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{C} \tag{1.45}$$

  and

$$a_{ij} + b_{ij} = c_{ij} \tag{1.46}$$

  Of course, we can add or subtract two matrices if and only if they have the same numbers of rows and columns. Moreover, matrix addition is commutative, i.e., $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$.

- Multiplication by a scalar:

  Multiplying a matrix $\boldsymbol{A}$ by a scalar k produces a new matrix $\boldsymbol{B}$ with the same number of rows and columns. Each element of $\boldsymbol{B}$ is obtained by multiplying the corresponding element of $\boldsymbol{A}$ by the scalar k:

$$kA = B \tag{1.47}$$

and

$$ka_{ij} = b_{ij} \tag{1.48}$$

- Matrix product:

  The product $\boldsymbol{AB}$ of two matrices is another matrix $\boldsymbol{C}$. This operation is possible if and only if the number of columns of the first matrix is equal to the number of rows of the second matrix. In general, the product of two matrices is not commutative:

$$\boldsymbol{AB} \neq \boldsymbol{BA} \tag{1.49}$$

  The product of two matrices in terms of their elements is:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{in}b_{nj} \tag{1.50}$$

- Transposed matrix

  By interchanging the rows and columns of a matrix $\boldsymbol{A}$, we obtain its transpose $\boldsymbol{A}^T$, so that: $a_{ij}^T = a_{ji}$, where the a_{ij}^T is the (i, j) entry of the transpose of $\boldsymbol{A}$, i.e.:

$$\boldsymbol{A}^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \tag{1.51}$$

1.4.4 The 2D Form of the Vector (Cross) Product

The vector product, or cross product, of two 3D vectors was defined in 1.3.5. This product exists only in three dimensions. However, in 2D geometry one is confronted frequently with the calculation of the cross product. To ease the solution of 2D geometric problems involving the cross product, we introduce below a *2D form of the cross product.*

Let $\boldsymbol{E}$ be an *orthogonal matrix* that rotates vectors in the plane through an angle of 90° counterclockwise (ccw), namely,

$$\boldsymbol{E} \equiv \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{1.52}$$

With this definition, we can readily prove that

$$\boldsymbol{E}^T\boldsymbol{E} = \boldsymbol{E}\boldsymbol{E}^T = \mathbf{1} \tag{1.53}$$

in which $\mathbf{1}$ is defined as the 2×2 identity matrix. Moreover, note that $\boldsymbol{E}$ is *skew-symmetric*, i.e., $\boldsymbol{E} = -\boldsymbol{E}^T$, and hence, $\boldsymbol{E}^2 = -\mathbf{1}$.

Also note that, given any vector $\boldsymbol{r} = \begin{bmatrix} x & y \end{bmatrix}^T$ in a plane Π, its image $\boldsymbol{r}$ under $\boldsymbol{E}$ is given by

$$\boldsymbol{Er} = \begin{bmatrix} -y \\ x \end{bmatrix} \tag{1.54}$$

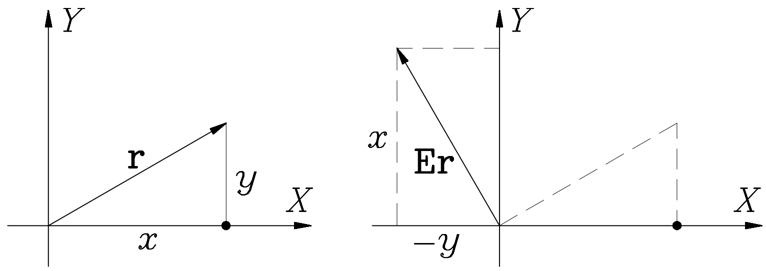


Figure 1.18: Vector $\boldsymbol{r}$ and its image under $\boldsymbol{E}$

as illustrated in Fig. 1.18.

Now, let us compute the cross product $\boldsymbol{a}\times\boldsymbol{b}$, where $\boldsymbol{a} = A\boldsymbol{k}$, and $\boldsymbol{k}$ is a unit vector normal to Π, pointing towards the viewer, $\boldsymbol{a}$ thus being a 3D vector normal to Π, of magnitude $|A|$. Moreover, we assume that $\boldsymbol{b}$ lies in Π, its Z-component thus vanishing. The cross product of interest thus takes the form

$$\boldsymbol{a}\times\boldsymbol{b} = \det\begin{bmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 0 & 0 & A \\ b_x & b_y & 0 \end{bmatrix} = -Ab_y\boldsymbol{i} + Ab_x\boldsymbol{j} \tag{1.55}$$

where we have recalled that the unit vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ are parallel to the X and Y axes, respectively. The 2-dimensional form of the foregoing product, then, becomes

$$(\boldsymbol{a}\times\boldsymbol{b})_{2D} = A\begin{bmatrix} -b_y \\ b_x \end{bmatrix} \equiv A\boldsymbol{E}\overline{\boldsymbol{b}} \tag{1.56}$$

where we have recalled eq.(1.54)and $\overline{\boldsymbol{b}}$ denotes the 2D version of $\boldsymbol{b}$, i.e.,

$$\overline{\boldsymbol{b}} = \begin{bmatrix} b_x \\ b_y \end{bmatrix}$$

Likewise, the cross product $\boldsymbol{b}\times\boldsymbol{c}$, for both $\boldsymbol{b}$ and $\boldsymbol{c}$ in the plane Π, is a vector perpendicular to this plane, of *signed* magnitude[1], $\|\boldsymbol{b}\|\|\boldsymbol{c}\|\sin(\boldsymbol{b},\,\boldsymbol{c})$, where $(\boldsymbol{b},\,\boldsymbol{c})$ denotes the angle between these two vectors, measured from $\boldsymbol{b}$ to $\boldsymbol{c}$.

Thus, if $\sin(\boldsymbol{b},\,\boldsymbol{c})$ is positive, the cross-product vector points in the direction of $\boldsymbol{k}$; otherwise, in the direction of $-\boldsymbol{k}$.

More concretely, let $\boldsymbol{b}$ be defined as before, $\boldsymbol{c}$ being defined, in turn, as:

$$\boldsymbol{c} \equiv \begin{bmatrix} c_x \\ c_y \\ 0 \end{bmatrix} \tag{1.57}$$

[1]The signed magnitude of a vector is a real number, positive, negative or zero, whose absolute value is identical to the magnitude of the vector.

Hence,

$$\boldsymbol{b} \times \boldsymbol{c} \quad l = \det \begin{bmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ b_x & b_y & 0 \\ c_x & c_y & 0 \end{bmatrix} = (b_x c_y - b_y c_x)\boldsymbol{k} \equiv C\boldsymbol{k} \tag{1.58}$$

where C is a real number that can be positive, negative, or even 0. Since we know the direction of $\boldsymbol{b} \times \boldsymbol{c}$, i.e., perpendicular to the plane Π, all we need is the quantity C above, which can be readily recognized as the dot product of the 2-dimensional vectors $\boldsymbol{E}\overline{\boldsymbol{b}}$, as given in eq.(1.56), and $\overline{\boldsymbol{c}}$, the 2D counterpart of $\boldsymbol{c}$, i.e.,

$$C = \boldsymbol{c}^T \boldsymbol{E} \boldsymbol{b} \equiv (\boldsymbol{E}\boldsymbol{b})^T \boldsymbol{c} = -\boldsymbol{b}^T \boldsymbol{E} \boldsymbol{c} \tag{1.59}$$

Therefore, the sign of C depends on whether the 3D cross product of eq.(1.58) points towards the reader or not. C vanishes, of course, if the two factors, $\overline{\boldsymbol{b}}$ and $\overline{\boldsymbol{c}}$, are parallel.

Now we introduce a practical application of the foregoing concepts in solving a recurrent problem of planar geometry:

Problem: *Given two lines $\mathcal{L}_1$ and $\mathcal{L}_2$, find the angle θ, for $0 \le \theta \le 2\pi$, that $\mathcal{L}_2$ makes with $\mathcal{L}_1$, while measuring θ ccw.*

Solution: Let $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ be 2D unit vectors indicating the direction of lines $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively. Obviously, $\cos\theta$ can be derived from the scalar product P1 of $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, namely,

$$\cos\theta = \boldsymbol{e}_1^T \boldsymbol{e}_2 \quad \text{or} \quad \boldsymbol{e}_1 \cdot \boldsymbol{e}_2 \tag{1.60}$$

However, the foregoing value does not determine uniquely θ, for, if P1 > 0, then θ may lie in either the first or the fourth quadrant, thus leaving us with an ambiguity. Ditto if P1 < 0, in which case θ may lie in either the second or the third quadrant.

To destroy the ambiguity, we need $\sin\theta$, which can be derived from the vector product of $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ when regarded as 3D vectors, as per eq. (1.30). In light of the 2D form of the vector product, however, we need not work out of the plane of the two given lines. Indeed, from eqs.(1.30),(1.58) and (1.59), we can write

$$\sin\theta = (\boldsymbol{E}\boldsymbol{e}_1)^T \boldsymbol{e}_2 \quad \text{or} \quad (\boldsymbol{E}\boldsymbol{e}_1) \cdot \boldsymbol{e}_2 \tag{1.61}$$

Let us call P2 the product appearing in the right-hand side of eq. (1.61).

Obviously, if P2 > 0, then θ may lie in either the first or the second quadrant, which leaves us with an ambiguity. Ditto if P2 < 0, in which case θ may lie in either the third or the fourth quadrant.

While each of P1 and P2 does not determine unequivocally angle θ individually, both do. In fact, we can draw the rules below:

1. If P1 > 0 and P2 > 0, then θ lies in the first quadrant;
2. If P1 < 0 and P2 > 0, then θ lies in the second quadrant;
3. If P1 < 0 and P2 < 0, then θ lies in the third quadrant;
4. If P1 > 0 and P2 < 0, then θ lies in the fourth quadrant.

1.4.5 Determinants

The *determinant* is a quantity associated with an arbitrary $n \times n$ square matrix (note that the number of rows and columns must be identical). A general definition of what a determinant actually represents is rather cumbersome; thankfully, we do not need it. We can define the determinant of a $n \times n$ matrix starting with the simplest case, i.e., $n = 2$, then $n = 3$, and hence, by induction, derive a *procedure* to compute the determinant for any arbitrary value of n.

As a matter of fact, the interest of the determinant is rather theoretical; its actual computation, which is extremely costly in terms of *floating-point operations*, or *flops*, is seldom needed. The relevance of the concept lies in that the value of the determinant indicates whether or not the matrix is *singular*, a case under which the matrix at hand cannot be *inverted*, a process useful for finding a solution to a linear system of n equations with n unknowns.

A 2×2 matrix $\boldsymbol{A}$ can be partitioned either column-wise or row-wise, as shown below:

$$\boldsymbol{A} \equiv \begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{c}^T \\ \boldsymbol{d}^T \end{bmatrix} \tag{1.62}$$

where $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$ are all 2-dimensional column vectors. Furthermore, we recall the definition of $\boldsymbol{E}$ as seen in Subsection 1.4.4. If the components of $\boldsymbol{a}$ and $\boldsymbol{b}$ are given as $\boldsymbol{a} = \begin{bmatrix} a_x & a_y \end{bmatrix}^T$ and $\boldsymbol{b} = \begin{bmatrix} b_x & b_y \end{bmatrix}^T$, the determinant of $\boldsymbol{A}$ is defined as

$$\det(\boldsymbol{A}) = \det \begin{bmatrix} a_x & b_x \\ a_y & b_y \end{bmatrix} = a_x b_y - a_y b_x \tag{1.63}$$

which can be readily cast in the form

$$\det(\boldsymbol{A}) = \begin{bmatrix} a_x & a_y \end{bmatrix} \begin{bmatrix} b_y \\ -b_x \end{bmatrix} \equiv \begin{bmatrix} a_x & a_y \end{bmatrix} \underbrace{\left(- \begin{bmatrix} -b_y \\ b_x \end{bmatrix} \right)}_{-\boldsymbol{E}\boldsymbol{b}} \tag{1.64}$$

the first array of the foregoing product being $\boldsymbol{a}^T$, the second $-\boldsymbol{E}\boldsymbol{b}$.

In summary, then:

$$\det(\boldsymbol{A}) = -\boldsymbol{a}^T \boldsymbol{E} \boldsymbol{b} = \boldsymbol{b}^T \boldsymbol{E} \boldsymbol{a} \tag{1.65}$$

A property of the determinant follows from its definition given in eq.(1.63):

$$\det(\boldsymbol{A}) = \det(\boldsymbol{A}^T) \tag{1.66}$$

If we recall the column-wise partitioning of $\boldsymbol{A}$, we can readily conclude that

$$\boldsymbol{A}^T = \begin{bmatrix} \boldsymbol{c} & \boldsymbol{d} \end{bmatrix} \tag{1.67}$$

whence,

$$\det(\boldsymbol{A}) = \det(\boldsymbol{A}^T) = -\boldsymbol{c}^T\boldsymbol{E}\boldsymbol{d} = \boldsymbol{d}^T\boldsymbol{E}\boldsymbol{c} \tag{1.68}$$

It should now be apparent that, if all the entities of a 2×2 matrix $\boldsymbol{A}$ are multiplied by the same scalar s, then its determinant is multiplied by s^2,

$$\det(s\boldsymbol{A}) = s^2\det(\boldsymbol{A}) \tag{1.69}$$

The determinant of a 3×3 matrix $\boldsymbol{A}$ is defined below. To this end, we partition $\boldsymbol{A}$ column-wise as

$$\boldsymbol{A} = \begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3\end{bmatrix} \tag{1.70}$$

The definition of $\det(\boldsymbol{A})$ can be expressed as:

$$\det(\boldsymbol{A}) = \boldsymbol{a_1} \times \boldsymbol{a_2} \cdot \boldsymbol{a_3} \tag{1.71}$$

The expression appearing in the right-hand side of the above equation is known as the *mixed product* or *triple product* of the three given vectors, and is sometimes represented as

$$[\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \boldsymbol{a}_3] = \boldsymbol{a}_1 \times \boldsymbol{a}_2 \cdot \boldsymbol{a}_3 \tag{1.72}$$

which is preserved under a *cyclic permutation*

Similar to relation (1.69), we have now, for a 3×3 matrix

$$\det(s\boldsymbol{A}) = s^3\det(\boldsymbol{A}) \tag{1.73}$$

Now we can define the determinant of a $n \times n$ matrix $\boldsymbol{A}$ for arbitrary n; this is done recursively, by defining this determinant as a *linear combination* of the determinants of $(n-1)\times(n-1)$ matrices. To this end, we denote the (i, j) entry of $\boldsymbol{A}$ as $a_{i,j}$. Moreover, the *minor* $M_{i,j}$ of entry $a_{i,j}$ is defined as the determinant of a $(n-1) \times (n-1)$ matrix $\boldsymbol{A}_{i,j}$, obtained from $\boldsymbol{A}$ upon deleting the ith column and the jth row, i.e.,

$$M_{i,j} = \det(\boldsymbol{A}_{i,j})$$

Further, the *cofactor* $C_{i,j}$ of entry $a_{i,j}$ of $\boldsymbol{A}$ is nothing but $M_{i,j}$ itself if $i+j$ is even; $-M_{i,j}$ if $i+j$ is odd. That is,

$$C_{i,j} = (-1)^{i+j}M_{i,j}$$

Thus,

- The determinant of $\boldsymbol{A}$ is defined in terms of the ith row as

$$\det(\boldsymbol{A}) = \sum_{j=1}^{n} a_{i,j}\, C_{i,j} \tag{1.74}$$

- If we recall property (1.66), an alternative definition of $\det(\boldsymbol{A})$ follows, in terms of the jth column:

$$\det(\boldsymbol{A}) = \sum_{i=1}^{n} a_{i,j}\, C_{i,j} \tag{1.75}$$

Actually, the definition of the determinant of a 3×3 matrix, eq.(1.71), follows from the counterpart definition of a 2×2 determinant, eq.(1.63), and the foregoing general definition of the determinant of an $n \times n$ matrix $\boldsymbol{A}$. Now we can state a generalization of relations(1.69) and (1.73):

The determinant of a $n \times n$ matrix $\boldsymbol{A}$ is homogeneous of degree n, i.e., if all the entries of $\boldsymbol{A}$ are multipliedby the same scalar s, *then* $\det(\boldsymbol{A})$ *becomes multiplied by s^n, i.e.,*

$$\det(s\boldsymbol{A}) = s^n \det(\boldsymbol{A}) \tag{1.76}$$

As stated earlier, the computation of the determinant of a $n \times n$ matrix from its definition, eq.(1.74) or, equivalently, eq.(1.75), is extremely costly. Indeed, from the foregoing discussion it is apparent that computing a 2×2 determinant requires two multiplications and one addition, or, roughly, two *flops* (one floating point operation, or flop, is made up of one addition and, concurrently, one multiplication). The computation of a 3×3 determinant requires the computation of three 2×2 determinants, which amounts to, roughly, six flops, but then, each of these determinants (cofactors) must be multiplied by its corresponding entry, these three products being finally added up, which brings about three more flops—give or take an addition operation. Hence, the computation of a 3×3 determinant requires $3 \times 2 + 3 = 3! + 3 = 9$ flops. Using induction to extrapolate a pattern, we can estimate that the computation of a $n \times n$ determinant consumes slightly over $n!$ flops. Now, the factorial grows extremely rapidly with n, which means that, even for moderately large values of n, $n!$ may lead to a prohibitively large number of flops.

As an example, let us consider a 30×30 matrix, which can frequently arise in various engineering applications. The number N of flops required to compute the determinant of such a matrix would be, as obtained with computer algebra,

$$N = 30! = 265252859812191058636308480000000$$

which is a pretty large number. To gain insight into the size of this number, let us assume that we have a Cray T90 supercomputer, capable of executing nearly 6×10^{10} flops/s. The time T such a computer would take can now be readily found as $T = N/6{\times}10^{10} = 4.420880997{\times}10^{21}$ s which is, again, a pretty large time interval. In order to have an idea of how big this time estimate is, let us compare it with the age of the universe, which lies somewhere between 10^{10} and 2×10^{10} years. In seconds, the lower estimate is 3.1536×10^{17} s. Hence, the ratio r of T to the lower estimate of the age of the universe is[2] $r = 14018.52168$, which indicates that a Cray T90 would need, roughly, 14 000 times the lower estimate of the age of the universe to compute such a determinant!

The good news is that streamlined methods are available to compute determinants, when such a computation is needed at all. One method, studied in courses on numerical analysis and applied linear algebra, relies on what is known as the *LU-decomposition* of matrix $\boldsymbol{A}$, under which this matrix is factored into the form $\boldsymbol{A} = \boldsymbol{LU}$. In this factoring, $\boldsymbol{L}$ is a lower-triangular matrix with only 1's on its diagonal and $\boldsymbol{U}$ is an upper-triangular matrix. A property of the

[2] Hawking, S.W., 1988, *A Brief History of Time. From the Big Bang to Black Holes*, Bantam Press, Toronto, New York, London, Sydney, Auckland, p. 108.

determinant states that the determinant of a product of matrices equals the product of the determinants of the individual matrix factors, and hence,

$$\det(\boldsymbol{A}) = \det(\boldsymbol{L})\det(\boldsymbol{U}) \tag{1.77}$$

By virtue of the structure of $\boldsymbol{L}$, we have $\det(\boldsymbol{L}) = 1$, and hence, $\det(\boldsymbol{A}) = \det(\boldsymbol{U})$. Moreover, given that $\boldsymbol{U}$ is upper-triangular, its determinant equals the product of its diagonal entries, which consumes only $n-1$ multiplications. The LU decomposition of a $n \times n$ matrix requires M_n multiplications and A_n additions[3]:

$$M_n = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} \tag{1.78}$$

$$A_n = \frac{n^3}{3} - \frac{n}{3} \tag{1.79}$$

For a 30×30 matrix, the foregoing figures amount to 9 155 multiplications and 8 990 additions, or roughly 9 000 flops. Any modern PC can execute approximately 10^8 flops/s, which means that the computation of a 30×30 determinant consumes about 100 μs, quite a short time interval when compared to the age of the universe!

Determinants of Block Matrices

Computing determinants of $n \times n$ matrices, for $n > 3$, can be achieved by resorting to the formulas available for *block-partitioned matrices*. For example, in Ch. 4, we may need to find the determinant of an affine transformation. In 3D, the affine transformation in question is given by a homogeneous 4×4 matrix, as introduced in Section 4.3. We thus consider here a $n \times n$ block matrix $\boldsymbol{P}$, where n is any natural number, defined by blocks, namely,

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \tag{1.80}$$

where we assume that all blocks are compatible, i.e., if $\boldsymbol{A}$ is of $p \times p$ and $\boldsymbol{D}$ is of $q \times q$, then $\boldsymbol{B}$ is of $p \times q$ and $\boldsymbol{C}$ of $q \times p$. We thus have implicitly assumed that $p + q = n$.

As an example, consider the 4×4 *homogeneous-transformation matrix* $\boldsymbol{T}$ of eq.(4.23), reproduced below for quick reference:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{M} & \boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{bmatrix} \tag{1.81}$$

In this case, the 3×3 matrix $\boldsymbol{M}$ represents a rotation, a reflection, or a scaling—these terms are explained in detail in Ch. 4—while the three-dimensional vector $\boldsymbol{t}$ represents a translation, $\boldsymbol{0}$ the three-dimensional zero vector and 1 is the real unity.

[3]Dahlquist, G. and Björck, Å., 1974, *Numerical Methods*, Prentice-Hall, Inc., Englewood Cliffs.

The formulas that allow the user to compute the determinant of the block-matrix $\boldsymbol{P}$ given in eq.(1.80) are displayed below[4]:

$$\det\left(\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}\right) = \det(\boldsymbol{A})\det(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})$$
$$= \det(\boldsymbol{D})\det(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})$$

Notice that any of the two foregoing formulas can be applied. However, the first formula requires that $\boldsymbol{A}^{-1}$ be invertible, while the second that $\boldsymbol{D}$ be so. In some cases, one of these two matrices is invertible, but not both. The user must choose judiciously which of the two formulas to apply. If none of $\boldsymbol{A}$ and $\boldsymbol{D}$ is invertible, to compute $\det(\boldsymbol{P})$, a reshuffling of the blocks may be needed, while taking into account that the sign of a determinant is preserved only under a *cyclic permutation* of either its columns or its rows.

As an example, we obtain the determinant of the 4×4 matrix of eq.(1.81), where we identify the blocks below:

$$\boldsymbol{A} = \boldsymbol{M}, \quad \boldsymbol{B} = \boldsymbol{t}, \quad \boldsymbol{C} = \boldsymbol{0}^T, \quad \boldsymbol{D} = 1 \tag{1.82a}$$

and hence, by application of the first of formulas (1.82), we have

$$\det(\boldsymbol{T}) = \det(\boldsymbol{M})\det(1 - \boldsymbol{0}^T\boldsymbol{M}^{-1}\boldsymbol{t})$$
$$= \det(\boldsymbol{M})(1) = \det(\boldsymbol{M})$$

which shows that, regardless of the value of vector $\boldsymbol{t}$, the determinant of the 4×4 *homogeneous-transformation matrix* $\boldsymbol{T}$ is always identical to that of $\boldsymbol{M}$.

1.4.6 Matrix Inversion

A $n \times n$ matrix whose determinant vanishes is termed *singular*; otherwise, the matrix is said to be *nonsingular*. Nonsingular matrices are sometimes referred to as *regular*.

Any $n \times n$ nonsingular matrix $\boldsymbol{A}$ has an associated *inverse*, denoted $\boldsymbol{A}^{-1}$, such that

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{1} \tag{1.83}$$

where $\boldsymbol{1}$ denotes the $n \times n$ identity matrix.

With the definition of cofactor introduced in Subsection 1.4.5, we can now define the *adjoint* $\mathrm{Adj}(\boldsymbol{A})$ of a $n \times n$ matrix $\boldsymbol{A}$ as the $n \times n$ matrix whose $(i,\, j)$ entry is the cofactor $C_{i,j}$ of $a_{i,j}$, namely,

$$[\mathrm{Adj}(\boldsymbol{A})]_{i,j} = C_{i,j} \tag{1.84}$$

Now, the inverse of $\boldsymbol{A}$ can be computed using the formula

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})}\mathrm{Adj}(\boldsymbol{A}) \tag{1.85}$$

[4]These formulas can be proven by various means; this proof not being pertinent to the course, it is left aside

In reality, the matrix inverse is seldom needed as such to perform computations in practical engineering problems, but it occurs frequently in analysis. Indeed, the matrix inverse occurs when solving a system of n linear equations in n unknowns. In these cases, the numerical procedure relies on the LU-decomposition of the matrix coefficient and the observation that a *triangular* system of equations admits a *recursive* solution involving only arithmetic operations. A system of equations is considered to be *upper-triangular* if the nth equation involves only the nth unknown, the $(n-1)$st equation only the nth and the $(n-1)$st unknowns, and so on, with the first equation involving all unknowns:

$$\begin{array}{rrrrl} a_{11}x_1+ & a_{12}x_2+ & a_{13}x_3+ & \ldots+ & a_{1n}x_n = b_1 \\ & a_{22}x_2+ & a_{23}x_3+ & \ldots+ & a_{2n}x_n = b_2 \\ & & & \ddots & \quad\vdots \quad = \vdots \\ & & & & a_{mn}x_n = b_n \end{array} \tag{1.86}$$

A *lower-triangular* system is defined likewise. Therefore, an upper-triangular system of linear equations can be readily solved *recursively* by *backward substitution*: Start by solving the nth equation for the nth unknown, thereby ending up with only $n-1$ unknowns left. The $(n-1)$st equation is next solved for the $(n-1)$st unknonwn, which leaves us with only $n-2$ unknowns to compute. At the beginning of the nth *recursion*, we are left with only one unknown, which can readily be solved for from the first equation.

We will not elaborate further on the solution of linear systems of equations for arbitrary values of n, but will rather focus on two special cases that can be handled *symbolically*, i.e. without a numerical procedure, but using formulas instead. Obviously, the simplest non-trivial cases occur when $n=2$ and $n=3$, as discussed below:

As the reader can readily verify, for a 2×2 matrix $\boldsymbol{A}$, partitioned as shown in eq.(1.62),

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})}\begin{bmatrix} \boldsymbol{b}^T \\ -\boldsymbol{a}^T \end{bmatrix}\boldsymbol{E} = \frac{1}{\det(\boldsymbol{A})}\boldsymbol{E}\begin{bmatrix} -\boldsymbol{d} & \boldsymbol{c} \end{bmatrix} \tag{1.87}$$

A quick verification involves only the computation of the product $\boldsymbol{A}\boldsymbol{A}^{-1}$, or $\boldsymbol{A}^{-1}\boldsymbol{A}$ for that matter, which should yield the 2×2 identity matrix.

Given a 3×3 matrix $\boldsymbol{A}$ partitioned as in eq.(1.70), its inverse may be evaluated in the form:

$$\boldsymbol{A}^{-1} = \frac{1}{\Delta}\begin{bmatrix} (\boldsymbol{a}_2\times\boldsymbol{a}_3)^T \\ (\boldsymbol{a}_3\times\boldsymbol{a}_1)^T \\ (\boldsymbol{a}_1\times\boldsymbol{a}_2)^T \end{bmatrix}, \quad \Delta \equiv \det(\boldsymbol{A}) = \boldsymbol{a}_1\times\boldsymbol{a}_2\cdot\boldsymbol{a}_3 \tag{1.88}$$

Again, the reader can verify the validity of the foregoing formula by straightforward computation of the product $\boldsymbol{A}\boldsymbol{A}^{-1}$ or, equivalently, of $\boldsymbol{A}^{-1}\boldsymbol{A}$.

Inverses of Block Matrices

Given the same block-matrix as in eq.(1.80), its inverse is given by

$$\boldsymbol{P}^{-1} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \\ \boldsymbol{Z} & \boldsymbol{U} \end{bmatrix} \tag{1.89}$$

with $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$, and $\boldsymbol{U}$ being, correspondingly, $p \times p$, $p \times q$, $q \times p$ and $q \times q$ blocks, whose values are given below:

$$\boldsymbol{X} = (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} \tag{1.90a}$$

$$\boldsymbol{U} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \tag{1.90b}$$

$$\boldsymbol{Y} = -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{U} \tag{1.90c}$$

$$\boldsymbol{Z} = -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{X} \tag{1.90d}$$

The validity of the foregoing formulas can be verified by straightforward computations: simply multiply matrix $\boldsymbol{P}$, as given in eq.(1.80), by $\boldsymbol{P}^{-1}$, as given in eq.(1.89). The product should yield the $n \times n$ identity matrix.

As an exercise, let us compute the inverse of the 4×4 homogeneous-transformation matrix $\boldsymbol{T}$ of eq.(1.81). In this case, we have, for $\boldsymbol{X}$ and $\boldsymbol{U}$,

$$\boldsymbol{X} = [\boldsymbol{M} - \boldsymbol{t}(1)^{-1}\mathbf{0}^T]^{-1}$$

$$\boldsymbol{U} = (1 - \mathbf{0}^T\boldsymbol{M}^{-1}\boldsymbol{t})^{-1}$$

In the above expressions, notice that the $\boldsymbol{D}$ block in matrix $\boldsymbol{T}$ is the real unity 1, which can be interpreted as the 1×1 "identity matrix," its inverse being the real unity itself. Hence, the product $\boldsymbol{t}(1)^{-1}\mathbf{0}^T$ in the brackets of the expression for $\boldsymbol{X}$ becomes

$$\boldsymbol{t}(1)^{-1}\mathbf{0}^T \equiv \boldsymbol{t}\mathbf{0}^T$$

which, as the reader can verify, is the 3×3 zero matrix $\boldsymbol{O}$, and hence,

$$\boldsymbol{X} = \boldsymbol{M}^{-1}$$

i.e., the inverse of $\boldsymbol{M}$. By the same token, the reader can verify that the product $\mathbf{0}^T\boldsymbol{M}^{-1}\boldsymbol{t}$ in the parenthesis of the expression for $\boldsymbol{U}$ reduces to the 1×1 "zero matrix," i.e., the real 0. As a consequence,

$$\boldsymbol{U} = 1^{-1} = 1$$

Therefore,

$$\boldsymbol{Y} = -\boldsymbol{M}^{-1}\boldsymbol{t}\boldsymbol{U} = -\boldsymbol{M}^{-1}\boldsymbol{t}$$

Moreover,

$$\boldsymbol{Z} = -\boldsymbol{D}^{-1}\boldsymbol{C}X = -1^{-1}\mathbf{0}^T\boldsymbol{M}^{-1} = \mathbf{0}^T$$

Finally, substituting all four expressions for $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$ and $\boldsymbol{U}$ in eq.(1.89), we obtain the desired inverse:

$$\boldsymbol{T}^{-1} = \begin{bmatrix} \boldsymbol{M}^{-1} & -\boldsymbol{M}^{-1}\boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{1.91}$$

That is, $\boldsymbol{T}$ and its inverse bear the same *gestalt*: the two lower blocks do not change, while the left-upper block becomes the inverse of the corresponding block in $\boldsymbol{T}$, the right-upper block becoming the negative of the product of the left-upper block in $\boldsymbol{T}^{-1}$ by the right-upper block in $\boldsymbol{T}$.

In summary, computing the inverse of a 4×4 homogeneous-transformation matrix, and of any 4×4 matrix for that matter, reduces to computing the inverse of a 3×3 matrix when the formulas (1.90a–d) are invoked. Since we have a formula for the inverse of a 3×3 matrix in eq.(1.84), it is straightforward to obtain a formula for the inverse of any particular 4×4 matrix. Finally, notice that the same formulas can be applied to compute the inverse of the 3×3 homogeneous-transformation matrix $\boldsymbol{T}$ introduced in Section 4.1.

Chapter 2

2D Objects

Geometric elements are categorized as: points, lines, surfaces, or solids. Lines, surfaces, and solids also have many subcategories. Points, lines, circles, and curves are the basic 2D geometric primitives, or *generators*, from which other, more complex geometric objects can be derived or algorithmically produced. For example, by taking a straight line and moving it in a certain way through a circular path, one can create a cylinder. This section defines, illustrates, and describes how to create points, lines, circles, and curves in the plane.

2.1 Points

A point is the simplest of the elementary geometric objects. Points are the basic building blocks for all other geometric objects, as shown in Fig. 2.1. Points are indispensable when we create computer graphic displays and geometric models.

A point is a geometric concept that has position but no dimensions. A point-position is defined by a set of real numbers, which are commonly referred to as coordinates. In the XY-plane, a point is represented by a pair of numbers, (x, y), where x and y are the signed distances from the Y- and the X-axes, respectively.

The location of point P may also be expressed as an array of numbers, known as the *position vector* $\boldsymbol{p}$ of P , namely

$$\boldsymbol{p} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{2.1}$$

Note: As stated previously, all vectors in this course are assumed to be *column arrays*. Moreover, *row arrays* may be obtained from column arrays by *transposition*, which is indicated by a right superscript T. We sometimes need to convert a column array to a row array for certain calculations.

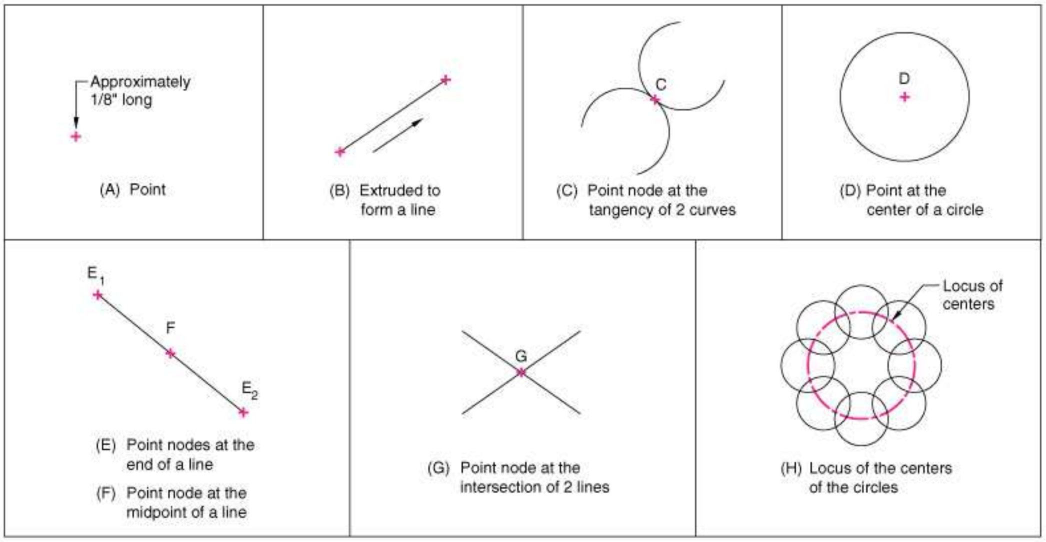


Figure 2.1: Examples and representation of points

2.2 Lines

Definition

A line is a geometric primitive that has length and direction, but not thickness; it is generated by a point moving in a constant direction.

Two elements of 2D geometry can define a line:

- two points
- a point and a vector parallel to the line
- a point and a vector perpendicular to the line

Algebraic representations of the line

Explicit Representation:

$$y = mx + p \tag{2.2}$$

This is the well-known slope-intercept form, where m and p are the slope and the Y-axis intercept, i.e., the intersection point of the line and the Y-axis.

Implicit Representation:

$$Ax + By + C = 0, \quad A^2 + B^2 > 0 \tag{2.3}$$

where A, B and C are constants. The explicit representation is expressed in the implicit form by substituting: $m = -A/B$ and $p = -C/B$, but this transformation requires $B \neq 0$, which is not always the case. Hence, the implicit representation is more general than its explicit counterpart.

Parametric Representation:

$$\begin{aligned} x &= au + b \\ y &= cu + d \end{aligned} \tag{2.4}$$

where u is the parameter, and a, b, c, d are constants.

2.2.1 Distance From a Point to a Line

Consider the line given by eq.(2.3). We want to compute the distance from a *given* point $Q(\xi,\, \eta)$ to the line. To this end, let us locate an arbitrary point $P_0(x_0,\, y_0)$ on the line. Since P_0 lies on the line, we cannot arbitrarily assign values to its coordinates. In fact, these must obey eq.(2.3):

$$Ax_0 + By_0 + C = 0 \tag{2.5}$$

whence we can solve for either y_0 in terms of x_0 or the other way around. *To reduce roundoff errors, it is advisable to solve for the unknown multiplied by the coefficient with the higher absolute value.* Once we have one unknown in terms of the other, all we need to do is assign an arbitrary value to the latter, which will thus produce a pair $(x_0,\, y_0)$ that complies with eq.(2.5). Upon assigning a second numerical value to the same unknown, we should be able to produce a second pair $(x_1,\, y_1)$ that also verifies eq.(2.5). We can now define two points P_0 and P_1 in the plane, of position vectors $\boldsymbol{p}_0$ and $\boldsymbol{p}_1$, given by

$$\boldsymbol{p}_0 = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, \quad \boldsymbol{p}_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \tag{2.6}$$

Next, we produce a unit vector $\boldsymbol{e}$ parallel to the line:

$$\boldsymbol{e} = \frac{\boldsymbol{p}_1 - \boldsymbol{p}_0}{\|\boldsymbol{p}_1 - \boldsymbol{p}_0\|} \tag{2.7}$$

Now, the unit normal $\boldsymbol{n}$ to the line can be most readily obtained by means of the $\boldsymbol{E}$ matrix introduced in eq.(1.52):

$$\boldsymbol{n} = \boldsymbol{E}\boldsymbol{e} \tag{2.8}$$

As the reader can readily verify, the distance d sought is simply

$$d = |\boldsymbol{n}^T(\boldsymbol{q} - \boldsymbol{p_0})| \tag{2.9}$$

where $\boldsymbol{q}$ is the position vector of $Q(\xi,\, \eta)$.

2.3 Planar Geometry and Polygons

2.3.1 Polygons

General Definition

A polygon is a multi-sided plane of any number of sides. If the sides of the polygon are equal in length and all its internal angles are equal, the polygon is known as a *regular polygon.*

A polygon with n edges is given by an ordered set of points $P_1, P_2, \ldots, P_n$; the said polygon has edge vectors $\boldsymbol{v}_i = \boldsymbol{p}_{i+1} - \boldsymbol{p}_i \; i = 1, \ldots, n$. which connect two neighbouring points to form the desired polygon, with $\boldsymbol{p}_i$ denoting the position vector of P_i.
Note: The number of vertices equals the number of edges.

Convexity

A classification of the polygons is based on *convexity*.

Convexity indicates that all points of the line segment defined by any two interior points or points on the perimeter, are either interior points or points on the perimeter.

One more interesting property: A regular n-sided convex polygon has a sum of interior angles I equal to:

$$I = (n - 2)\pi \tag{2.10}$$

Types of polygons

There exist many different types of polygons, but regular polygons are defined as being:

equilateral, which means that all sides are of equal length; and

equiangular, which means that all interior angles at the vertices are equal.

The polygons that exhibit these characteristics are also referred to as *n-gons*, where n indicates the number of edges. Thus, an equilateral triangle would be a 3-gon (or trigon), a square would be a 4-gon (or tetragon), and so on.

2.3.2 Regular Polygons

Among the different types of polygons that exist, the most useful is the *regular* polygon, which is equilateral and equiangular.

Regular polygons are grouped by their number of sides and are illustrated in Fig. 2.2.

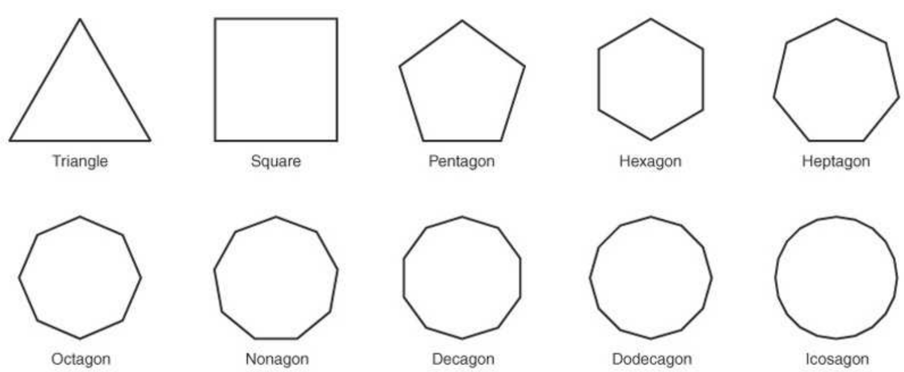


Figure 2.2: Regular polygons

2.4 Quadratic Curves: Conics

Quadratic curves, or conics, are the simplest of all 2D curves. Conics are used extensively in computer graphics and geometric modelling.

In the most general sense, conics are curves formed by the intersection of a plane with a right circular cone. The relative inclination of the plane with respect to the cone determines the conic produced: circle, ellipse, parabola, or hyperbola, as shown in Fig 2.3.

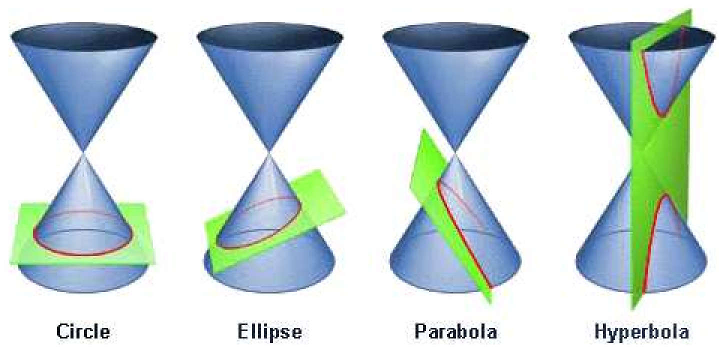


Figure 2.3: Conics

Conics are commonly described in implicit form by the quadratic equation

$$Ax^2 + By^2 + 2Cxy + 2Dx + 2Ey + F = 0 \tag{2.11}$$

where (x, y) are the coordinates of an arbitrary point of the curve, and A, B, C, D, E, F are the coefficients characterizing the type of conic produced.

In array form, and using homogeneous coordinates, this equation can be written as:

$$\boldsymbol{p}^T \boldsymbol{R} \boldsymbol{p} = 0 \tag{2.12}$$

where

$$\boldsymbol{p} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \boldsymbol{R} = \begin{bmatrix} A & C & D \\ C & B & E \\ D & E & F \end{bmatrix} \tag{2.13}$$

2.4.1 Circles

A circle is a geometric primitive, whose points are equidistant from one point, the centre of the circle. A circle is created when a plane passes through a right circular cone or cylinder, and is perpendicular to the axis of the cone (or cylinder, as the case may be), as shown in Fig. 2.3. Circles and their arcs are used extensively in engineering design, in particular for the design of mechanical parts.

Algebraic representation of the circle

Implicit Representation:

$$(x - x_1)^2 + (y - y_1)^2 = r^2 \tag{2.14}$$

If the centre is located at the origin (0, 0), the above equation simplifies to:

$$x^2 + y^2 = r^2 \tag{2.15}$$

Parametric Representation (of a circle centred at the origin):

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \end{aligned} \tag{2.16}$$

where θ is the angle between the radius to the point (x, y) and the x-axis. If the centre is not the origin but a point of coordinates (x_1, y_1), then the parametric form becomes:

$$\begin{aligned} x &= x_1 + r\cos\theta \\ y &= y_1 + r\sin\theta \end{aligned} \tag{2.17}$$

Finally, we obtain the array representation of the circle in the form

$$\boldsymbol{p}^T \boldsymbol{R}\boldsymbol{p} = \begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & D \\ 0 & 1 & E \\ D & E & F \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \tag{2.18}$$

2.4.2 Ellipses

Definition

An ellipse is a curve created when a plane intersects a right circular cone, at an acute angle with the cone axis greater than the acute angle between the axis and the cone elements, as shown in Fig. 2.4.

An ellipse can be defined alternatively as the locus of all points in a plane for which the sum of the distances from two fixed points F_1, F_2 (the foci) in the plane is constant: $\overline{PF}_1 + \overline{PF}_2 = \text{const.}$

On a practical note, an ellipse can be quickly constructed using a pencil attached to two strings, which are in turn attached to the two foci of the desired ellipse.

The **major axis** of an ellipse is the longest line segment included in the ellipse and passes through both foci.

The **minor axis** is the perpendicular bisector of the major axis.

A circle viewed at an angle other than 90° (normal), appears as an ellipse due to perspective, as we can see in Fig. 2.5.

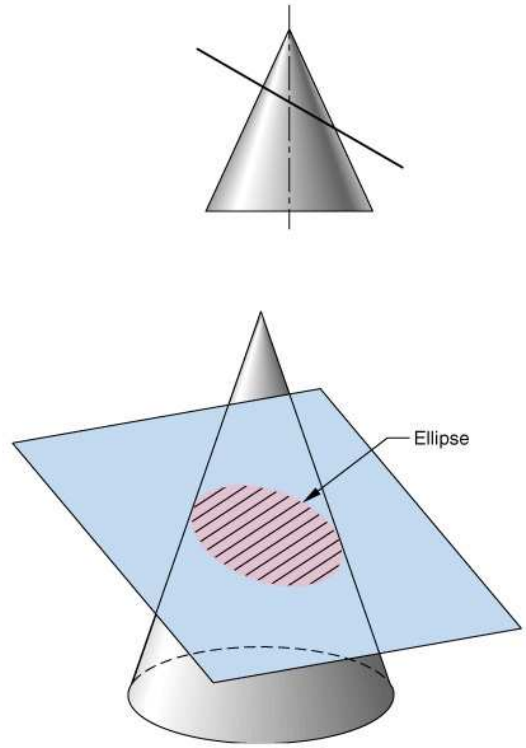


Figure 2.4: The ellipse

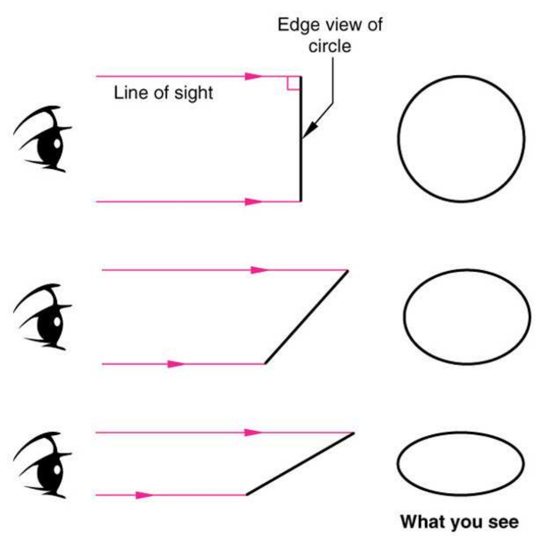


Figure 2.5: Circles viewed as ellipses

The ellipse possesses a reflective property similar to that of a parabola (see Section 2.4.3): light or sound emanating from one focus is reflected to the other, a property useful in the design of some types of optical and auditory equipment. Whispering galleries, such as the Rotunda in the Capitol Building in Washington, D.C., and the Mormon Tabernacle in Salt Lake City, Utah, were designed using elliptical ceilings.

In a whispering gallery like the one shown in Fig. 2.6, sound emanating from one focus is clearly audible at the other focus.

Figure 2.6: An ellipse application: The whispering gallery in the Great Rotunda, Washington, D.C.

Examples of ellipses from the real world can be observed if:

- A glass of water in a cylindrical glass is tilted; the free surface of the liquid will acquire an elliptical shape, as seen in Fig. 2.7.
- We observe the path that planets trace out around our sun. In the 17th century, Johannes Kepler discovered that each planet travels around the sun in an elliptical orbit with the sun at one of its foci, as illustrated in Fig. 2.8.
- We can also cite classical atomic theory: the electrons of an atom move in an approximately elliptical orbit, with the nucleus at one focus, as shown in Fig. 2.9.

Algebraic representation of the ellipse

Implicit Representation (of an ellipse centered at the origin):

The implicit representation of an ellipse *in canonical form* is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{2.19}$$

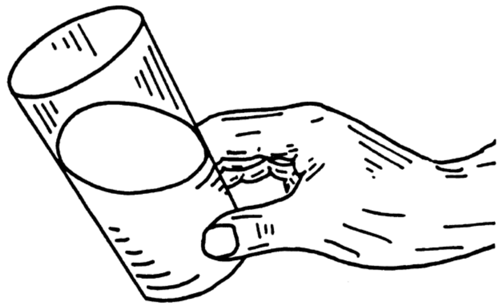


Figure 2.7: Tilting a glass of water

Figure 2.8: Orbit of the planets

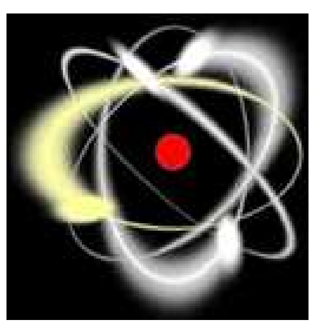


Figure 2.9: Orbit of an electron

where the axes of this ellipse are assumed to coincide with the coordinate axes. The constants a and b indicate the axis lengths.

Note: When the two axes are of the same length, the ellipse reduces to a circle.

Parametric Representation:

$$\begin{aligned} x &= a\cos\theta \\ y &= b\sin\theta \end{aligned} \tag{2.20}$$

where θ is the parameter, $\boldsymbol{a}$ and $\boldsymbol{b}$ are the axis lengths, and the axes of the ellipse coincide with the coordinate axes.

Finally, we obtain the array form of the general ellipse equation in terms of the homogeneous coordinates of one of its points:

$$\boldsymbol{p}^T\boldsymbol{R}\boldsymbol{p} = \begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} A & C & D \\ C & B & E \\ D & E & F \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \tag{2.21}$$

where A and B have the same sign.

2.4.3 Parabolas

Definition

A parabola is the curve created when a right circular cone is cut by a plane parallel to the element of the cone, as we see in Fig. 2.10.

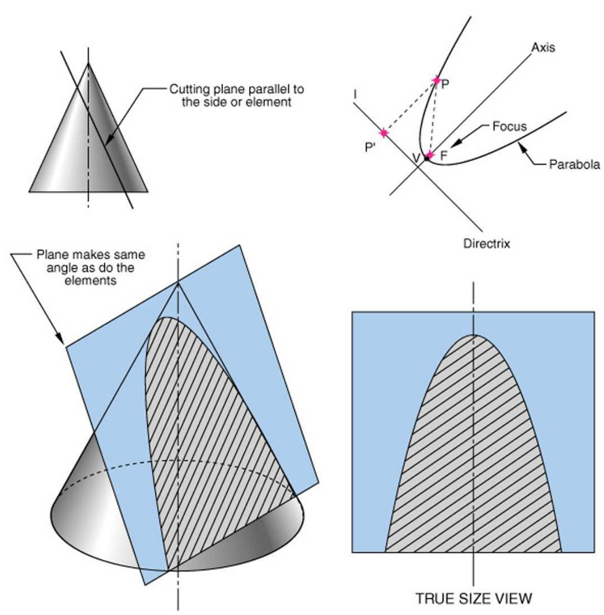


Figure 2.10: The parabola viewed as a conic section

A parabola can be defined alternatively as the locus of points in a plane that are equidistant from a given fixed point, called the *focus*, and a fixed line, called the *directrix*.

Parabolas are used in the design of mirrors for telescopes, reflective mirrors for lights, cams for uniform acceleration, weightless flight trajectories, antennae for radar systems, arches for bridges, and field microphones commonly seen on the sidelines of football games.

Parabolas are quite useful in the design of technological equipment due to a unique reflective property (see Fig. 2.11): Rays that originate at the focus of a parabola are reflected out of the parabola parallel to the axis. Conversely, rays entering the parabola parallel to the axis are reflected to the focus.

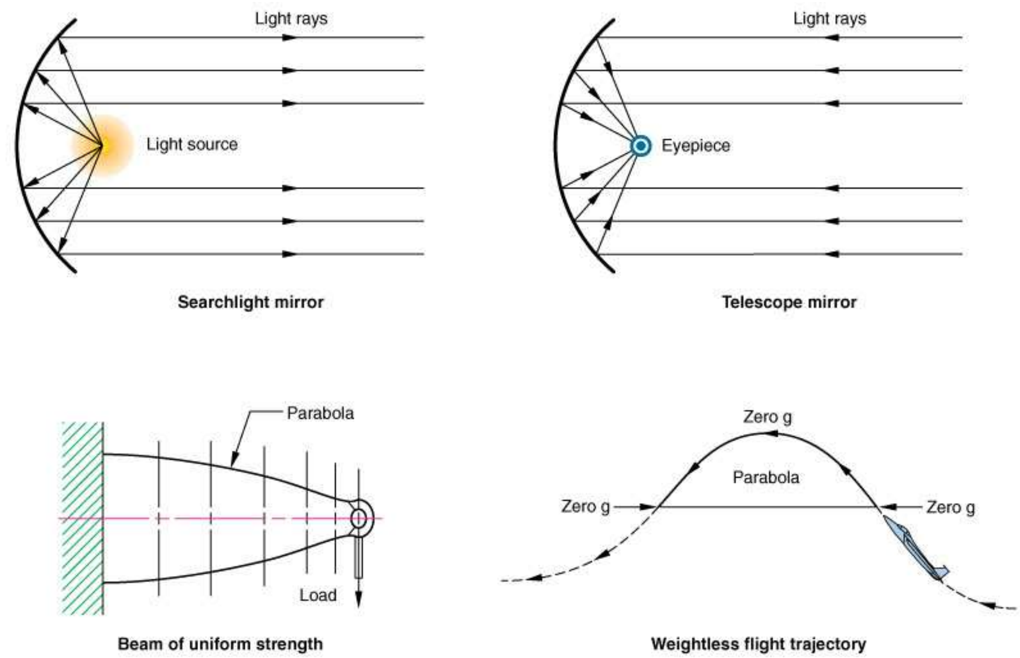


Figure 2.11: Engineering applications using the parabola properties

Parabolas can also be found in many other places:

- One of nature's best known approximations to parabolas is the path taken by a body projected upward and obliquely to the pull of gravity, as in the parabolic trajectory of a golf ball, as shown in Fig. 2.12.
- A parabolic trajectory is also exhibited by water emanating from a spout at a drinking fountain. Each molecule of water follows a parabolic path, thus providing a picture of the curve, as shown in Fig. 2.13.
- In the design of communications equipment, antennas are often used to collect radio waves and light from a variety of sources. The parabolic nature of the antenna allows it to collect and focus the signal at the focal point.

Algebraic Representation of the Parabola

Explicit Representation:

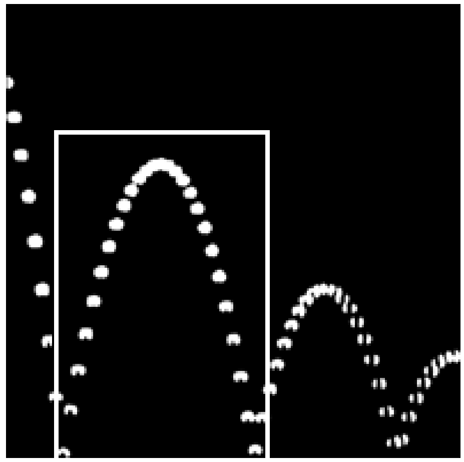


Figure 2.12: Trajectory of a golf ball

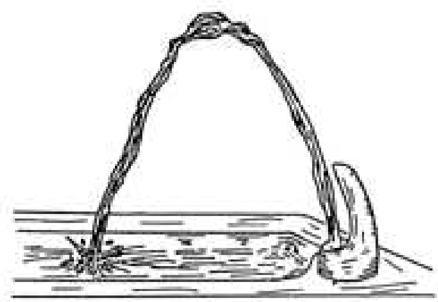


Figure 2.13: Trajectory of water ejected from a waterspout

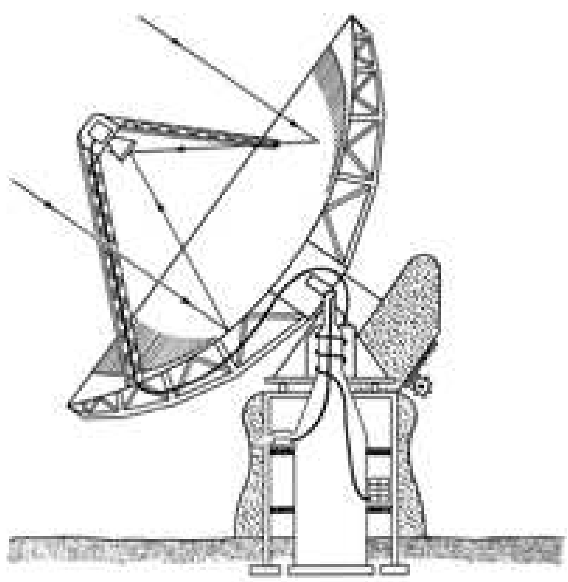


Figure 2.14: Antenna

The canonical form of a parabola passing through the origin and with focal axis coinciding with the Y-axis is given below as an explicit equation of y in terms of x:

$$x^2 = 4ay \tag{2.22}$$

where the axis of symmetry of this parabola is assumed to be the Y-axis, whereas a is a constant.

A parabola may also be represented by the equation:

$$y = ax^2 + bx + c \tag{2.23}$$

where a, b and c are constants. This is the most familiar representation of the parabola.

Parametric Representation:

The canonical form of eq.(2.22) is transformed into parametric form, using the parameter t:

$$\begin{aligned} x &= 2at \\ y &= at^2 \end{aligned} \tag{2.24}$$

Finally, the array form of the parabola is given by

$$\boldsymbol{p}^T\boldsymbol{R}\boldsymbol{p} = \begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} A & C & D \\ C & B & E \\ D & E & F \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \tag{2.25}$$

where $AB - C^2 = 0$.

2.4.4 Hyperbolas

Definition

The hyperbola is a curve of intersection created when a right circular cone is cut by a plane that makes a smaller angle with the axis than with the cone elements.

A hyperbola can be defined alternatively as the locus of all points in a plane whose distances from two fixed points, called the foci (lying in the plane as well), have a constant difference, as illustrated in Fig. 2.16.

Hyperbolas can also be found in many places:

- When alpha particles are shot towards the nucleus of an atom, they are repulsed away from the nucleus along hyperbolic paths.
- In astronomy, a comet that does not return to the sun follows a hyperbolic path.
- Hyperbolas are used in reflecting telescopes.
- A hyperbola revolving around its axis forms a surface called a hyperboloid. The cooling tower of a steam power plant has the shape of a hyperboloid, as does the architecture of the James McDonnell Planetarium of the St. Louis Science Center that we can see in Fig. 2.17.

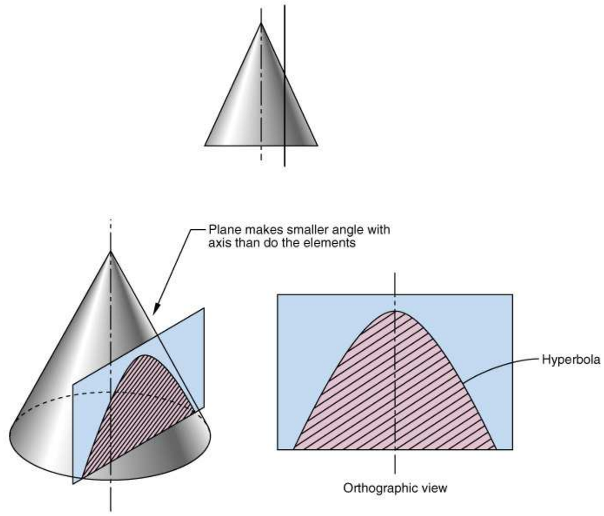


Figure 2.15: The hyperbola as a conic section

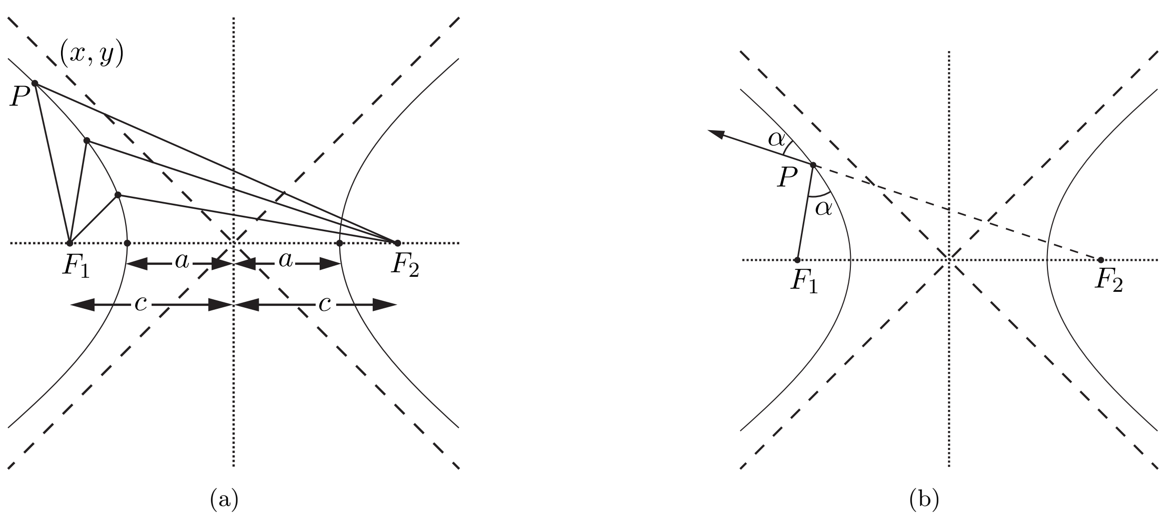


Figure 2.16: The hyperbola defined as the locus of all points P obeying the property $\overline{PF}_2 - \overline{PF}_1 = \text{constant}$, where F_1 and F_2 are the *foci*

Figure 2.17: The hyperbola in architecture: The James S. McDonnell Planetarium of the St. Louis Science Center

Algebraic representation of the hyperbola

Implicit Representation:

The canonical representation of the hyperbola:

$$\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1 \tag{2.26}$$

where the hyperbola axes are assumed to coincide with the coordinate axes. The definitions of major and minor axis for the hyperbola are identical to those given for the ellipse. The Y-axis intersects the curve at two points $(0,\, b)$ and $(0,\, -b)$, but the X-axis does not intersect the curve at all.

Parametric Equation (of the hyperbola with axes coincident with the coordinates axes):

$$\begin{aligned} x &= a\sinh\theta \\ y &= b\cosh\theta \end{aligned} \tag{2.27}$$

where $\sinh\theta$ and $\cosh\theta$ are the *hyperbolic trigonometric functions*, defined as

$$\cosh x = \frac{\exp(x) + \exp(-x)}{2}, \quad \sinh x = \frac{\exp(x) - \exp(-x)}{2} \tag{2.28}$$

where the axes of the hyperbola are assumed to coincide with the coordinates axes.

Furthermore, notice that the canonical form of the equation of the hyperbola can be cast in the form

$$(ay + bx)(ay - bx) = a^2b^2 \tag{2.29}$$

which is obviously violated by all points lying on the lines

$$\mathcal{L}_1:\ ay + bx = 0, \quad \mathcal{L}_2:\ ay - bx = 0 \tag{2.30}$$

Lines $\mathcal{L}_1$ and $\mathcal{L}_2$ are called the *asymptotes*—Greek: *a*, negation; *symptotos*, falling together—of the hyperbola.

Finally, we obtain the array form of the generalized hyperbola:

$$\boldsymbol{p}^T \boldsymbol{R} \boldsymbol{p} = \begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} A & C & D \\ C & B & E \\ D & E & F \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \tag{2.31}$$

where A and B have opposite signs.

Summary

In general, all types of conics have many engineering applications and can be utilized in combination with one another to create intricate machines and tools; the telescope in Fig. 2.18 uses mirrors to create the ellipses and hyperbolas. The parabola at the bottom of the apparatus was created by rotating liquid mercury in a cylindrical reservoir.

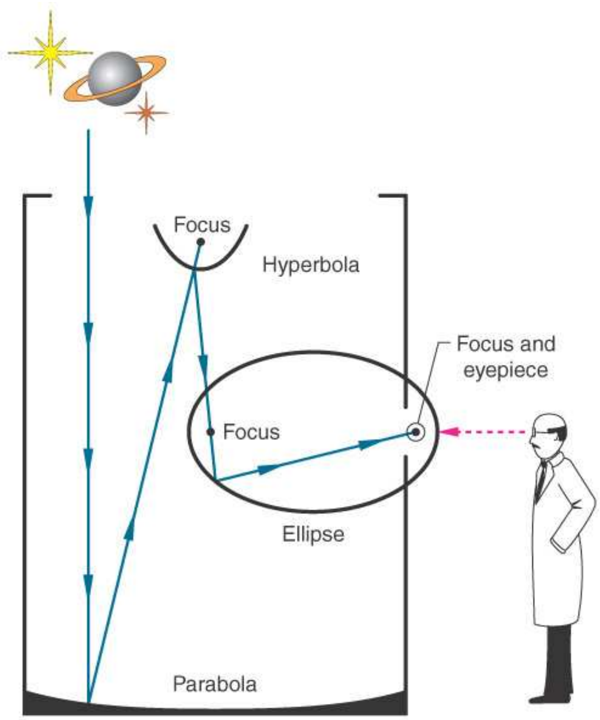


Figure 2.18: Example: application of conics to construct a telescope

Using the generalized implicit form of the general equation for conics,

$$Ax^2 + By^2 + 2Cxy + 2Dx + 2Ey + F = 0 \tag{2.32}$$

We can identify the type of conic simply by observing the sign of the *discriminant* of the general second-degree polynomial:

$$\Delta = AB - C^2 \tag{2.33}$$

which is the determinant of the top-left 2×2 block of the 3×3 matrix in eq.(2.13). Thus, we can identify the type of conic:

If $\Delta > 0$, the general equation represents an ellipse.

If $\Delta = 0$, the general equation represents a parabola.

If $\Delta < 0$, the general equation represents a hyperbola.

2.5 Higher-Order Algebraic Curves

Let us consider a curve $\mathcal{C}$ described by an implicit equation $f(x, y) = 0$, where function $f(x, y)$ involves products $x^i y^j$, for i and j any real integers. Such a curve, described by a *bivariate polynomial* in x and y, is termed an *algebraic curve*. The order, or degree, $d_{\mathcal{C}}$ of an algebraic curve $\mathcal{C}$ is defined as

$$d_{\mathcal{C}} \equiv \max_{i,j}\{\, i + j \,\} \tag{2.34}$$

The conics are, thus, second-order curves. Any algebraic curve with $d_{\mathcal{C}} > 2$ will be termed in this course a *higher-order curve*. Now we have the result below:

Fact 1 An algebraic curve of degree d intersects a line at d points at most.

Proof: The proof is straightforward. Consider the line $\mathcal{L}$ given by

$$\mathcal{L}: \quad Ax + By + C = 0 \tag{2.35}$$

Under the assumption that $B \neq 0$—if B turns out to vanish, then we can solve for x as $-C/A$, a constant value, and proceed in a slightly different, although simpler way—we can solve for y in terms of x, namely,

$$y = -\frac{A}{B}x - \frac{C}{B}$$

When this expression is substituted into $f(x, y) = 0$, a *monovariate polynomial* equation $P(x) = 0$, of degree d, is obtained. Now, the equation thus resulting has exactly d roots, whether real or complex, with complex roots occurring in conjugate pairs. Each such real root thus defines one intersection of $\mathcal{L}$ with $\mathcal{C}$, thereby proving that the line intersects the curve at d points at most.

A special class of algebraic curves that finds ample applications in design are *Lamé curves*, thus named after the French mathematician Gabriel Lamé (1795–1870), who first proposed them. These are m-order curves, endowed with interesting properties, which take the simple forms

$$f(x, y) = x^m + y^m - 1 = 0 \tag{2.36}$$

These curves are plotted in Fig. 2.19 for $m = 2, \ldots, 7$. Properties of these curves are given below:

- Lamé curves of even degree are closed and symmetric with respect to the x and y axes;

- Lamé curves of odd degree are open and symmetric with respect to a line passing through the origin and making an angle of 45° with the X-axis; and
- the curvature of the Lamé curves vanishes at the intersections with the coordinate axes, except for the case $m = 2$, in which case the curvature is constantly equal to unity.

While the Lamé curves defined above are *normalized*, in that the coefficients of x and y are unity, scalings in the directions of the coordinate axes are possible by means of an *affine transformation*, as outlined in Subsection 4.1.1.

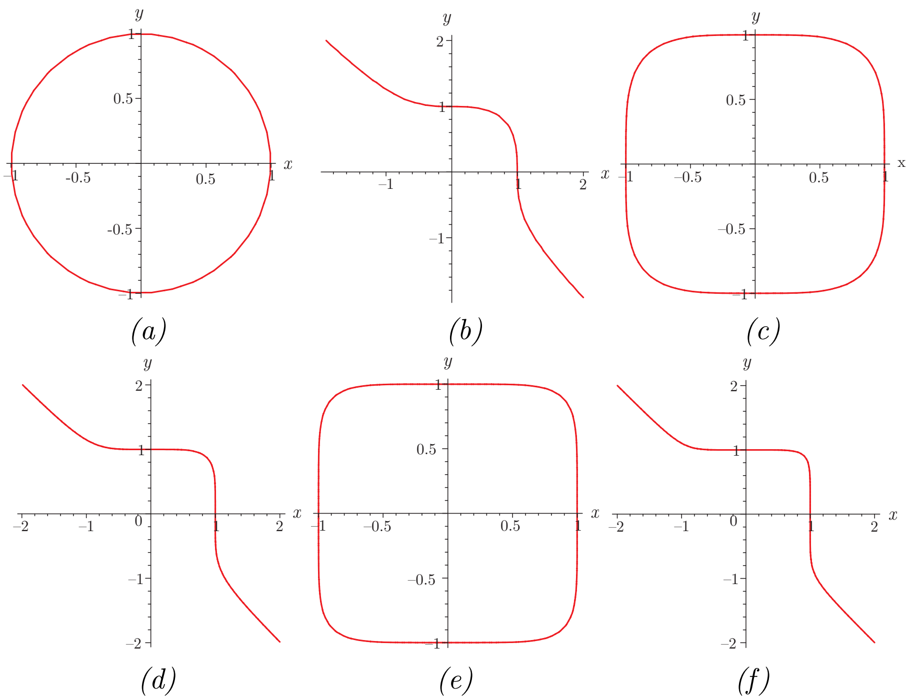


Figure 2.19: Plots of the Lamé curves for $m = 2, \ldots, 7$

It is noteworthy that the Lamé curves for $m = 3$, 5 and 7 in Fig. 2.19 exhibit an open-ended shape, i.e., these curves extend infinitely towards the second and the fourth quadrants. This feature, then, indicates the existence of an asymptote of the curve, similar to those of the hyperbola, as in Subsection 2.4.4. Similar to the case of the hyperbola, we can find the asymptote of the cubic Lamé curve upon first writing eq.(2.36), for $m = 3$, in the form

$$x^3 + y^3 = 1$$

Next, we factor the left-hand side of the above equation into a linear and a quadratic factor, namely,

$$(x + y)(x^2 - xy + y^2) = 1 \tag{2.37}$$

Apparently, the set of points on the line $x + y = 0$ never touches the curve, and hence, this line is the asymptote of the curve. As a matter of fact, all odd-order Lamé curves have the same asymptote.

The general (implicit) equation representing a cubic curve takes the form

$$
\begin{aligned}
f(x,\, y) \quad \equiv \quad & A_{30}x^3 + A_{21}x^2y + A_{12}xy^2 + A_{03}y^3 + A_{20}x^2 + A_{11}xy + A_{02}y^2 \\
& + A_{10}x + A_{01}y + A_{00} = 0
\end{aligned} \tag{2.38}
$$

Finding the asymptote of a general cubic may be more challenging, for this requires finding two factors, one linear and one quadratic, of $f(x,\, y)$, which is not a simple task. More general procedures are available for computing asymptotes of curves, but these fall beyond the scope of this course, and will hence not be pursued.

Moreover, non-algebraic curves are curves described by the implicit equation $f(x,\, y) = 0$, in which $f(x,\, y)$ is not a polynomial function. In this case, the number of intersections of the curve with a line may be infinite. Examples of nonalgebraic curves abound, e.g., the *logarithmic spiral*, the *cycloid*, the *circle-involute*, and so on. Examples of these curves are included in Fig. 2.20. The circle-involute, or simply the *involute*, is the curve used to produce gears.

The parametric equation of the *logarithmic spiral*, in polar coordinates, is

$$
r = a\exp(b\theta), \tag{2.39}
$$

while the parametric equations of the *cycloid* are

$$
\begin{aligned}
x_c(t) &= a(t - \sin t) \\
y_c(t) &= a(1 - \cos t),
\end{aligned} \tag{2.40}
$$

and those of the *circle-involute* are

$$
\begin{aligned}
x_i(t) &= a(\cos t + t\sin t) \\
y_i(t) &= a(\sin t - t\cos t)
\end{aligned} \tag{2.41}
$$

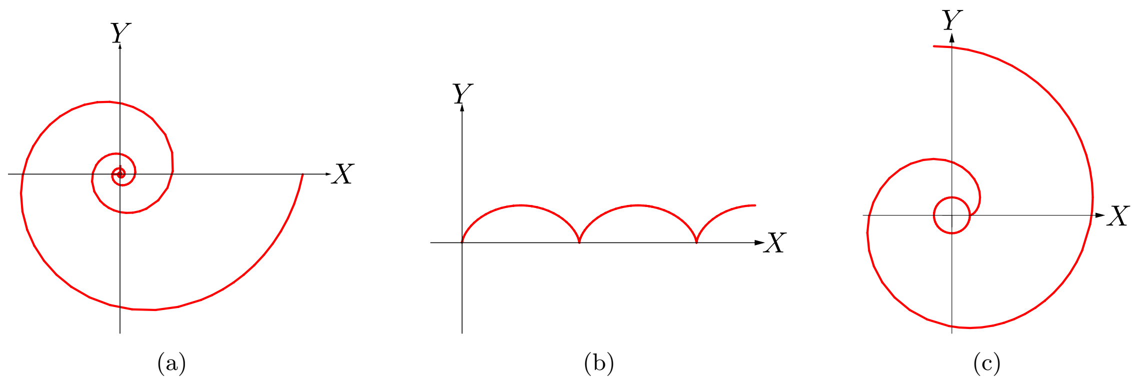


Figure 2.20: Examples of non-algebraic curves: (a) the logarithmic spiral; (b) the cycloid; (c) and the circle-involute

2.6 Free-form curves

Sometimes, design applications call for curves that cannot be represented, either in full or piecewise, by simple implicit functions of the form $f(x, y) = 0$. These curves are called *free-form curves*. The automobile industry uses many free-form curves in the design of the body,

Figure 2.21: Example of application of free-form curves

as shown in Fig. 2.21.

The *spline* is one of the most frequently used curves in the aircraft and ship-building industries. The cross-section of an airplane wing or a ship hull is a spline curve. In addition, spline curves are commonly used to define the path of motion in computer animation. For CAD systems, three types of free-form curves have been developed: splines, Bézier curves, and B-spline curves, which we can see in Fig. 2.22.

Bézier curves were invented simultaneously by Paul de Casteljau at Citroën and Pierre E. Bézier at Renault around the late 50s and early 60s. However, Bézier was able to publish his work in several journals, thus bestowing his name on the newly created family of curves.

These curves can be described by sets of parametric equations, in which the x and y coordinates of the *control points* are computed functions of a third variable, called a parameter.

The topic of free-form curves is rather advanced, for which reason it is not pursued in this course.

2.7 Curve-Blending

In geometry construction, we come frequently across the problem of *curve blending*, or *blending*, for brevity. For example, the spline displayed in Fig. 2.22(a) shows three intermediate points where four algebraic curves meet pairwise. At each of these points, the two curves are *forced* to share the point in question, which is termed G^0-continuity, with G standing for *geometric*, as opposed to *algebraic* continuity, which is represented with C. The common point is termed a *blending point*. Moreover, with reference to the same figure, the two *blent curves* share not only one common point, but also one common tangent, which is termed G^1-continuity. If, furthermore, the two blent curves are forced to share the same *centre of curvature*, and hence, the same curvature, then, we speak of G^2-continuity. Higher-order continuity is needed in some applications. In this course, however, we will not consider such special applications.

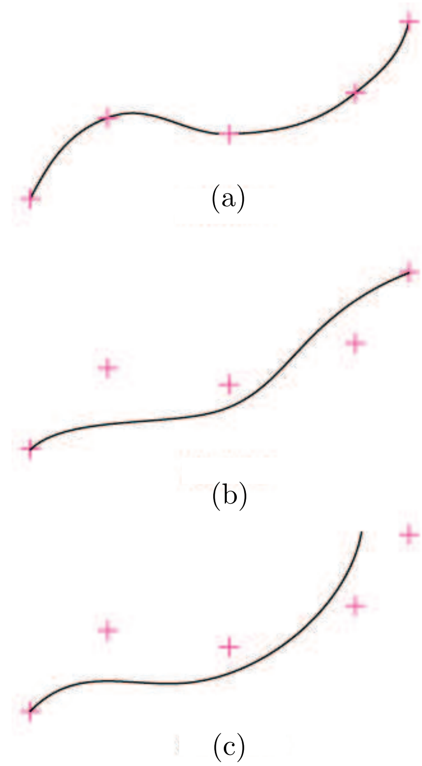


Figure 2.22: Free-form curves: (a) spline; (b) Bézier; (c) B-spline

Chapter 3

3D Objects

3.1 Points, Lines and Planes in Space

A point P is defined in three dimensions by its three Cartesian coordinates (x, y, z), and represented by its position vector $\boldsymbol{p}$:

$$\boldsymbol{p} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{3.1}$$

3.1.1 Planes

A plane is the locus of points equidistant from two fixed points. The resulting plane is the perpendicular bisector of the line joining the two points. This definition is known as *demonstrative* or *constructive*.

Computer-graphics and geometric-modelling require a more quantitative definition. We derive below the implicit equation of the plane.

The relation between the constructive definition and the analytic-geometric definition sought can be readily derived. Let P_1 and P_2 be the two points in question, their position vectors being $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, respectively. Equating the distances, or their squares for that matter, of any point P of the plane to P_1 and P_2, we obtain

$$||\boldsymbol{p}_1 - \boldsymbol{p}||^2 = ||\boldsymbol{p}_2 - \boldsymbol{p}||^2$$

Each side of the above equation bears striking similarities with the square of a binomial. It is left as an exercise to the reader to prove that the sides of that equation expand as

$$||\boldsymbol{p}_1||^2 - 2\boldsymbol{p}_1^T\boldsymbol{p} + ||\boldsymbol{p}||^2 = ||\boldsymbol{p}_2||^2 - 2\boldsymbol{p}_2^T\boldsymbol{p} + ||\boldsymbol{p}||^2$$

which readily reduces to

$$(\boldsymbol{p}_2 - \boldsymbol{p}_1)^T\boldsymbol{p} + \frac{1}{2}(||\boldsymbol{p}_1||^2 - ||\boldsymbol{p}_2||^2) = 0 \tag{3.2}$$

Now, let

$$\boldsymbol{p}_2 - \boldsymbol{p}_1 \equiv \begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{bmatrix} = \begin{bmatrix} A \\ B \\ C \end{bmatrix}, \quad D = \frac{1}{2}(||\boldsymbol{p}_1||^2 - ||\boldsymbol{p}_2||^2) \tag{3.3a}$$

whence the implicit equation sought becomes

$$Ax + By + Cz + D = 0 \tag{3.3b}$$

which is a linear equation in x, y and z. If the coordinates of any point satisfy eq.(3.3b), then the point lies in the plane.

Notice the similarity between the implicit equation of a line in 2D, eq. (2.3), and that of a plane, as derived above.

3.1.2 Lines in Space

In three-dimensional space, a line is defined by a base point A of position vector $\boldsymbol{a}$ and a direction vector $\boldsymbol{b}$, which gives the direction of the line. Therefore, the vector equation of a line is: $\boldsymbol{p} = \boldsymbol{a} + u\boldsymbol{b}$ where $\boldsymbol{p}$ is the position vector of an arbitrary point P of the line and u is a real parameter. Unless otherwise stated, the direction vector is assumed to be of unit magnitude.

The line may also be represented in the form of three linear parametric equations, one for each coordinate:

$$\begin{aligned} x &= a_x + b_x u \\ y &= a_y + b_y u \\ z &= a_z + b_z u \end{aligned} \tag{3.4}$$

where x, y, z are the coordinates of an arbitrary point of the line, or the components of vector $\boldsymbol{p}$; hence, these coordinates are the dependent variables. The set of equations in (3.4) generates a set of coordinates for each value of the parameter u. The coefficients $a_x, a_y, a_z, b_x, b_y, b_z$ are unique and constant for any given line. Each of these two triplets is the set of components of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ above.

Alternatively, a line can be represented as the intersection of two planes. Each plane equation takes the form (3.3b), and hence, the two planes are represented by

$$A_1 x + B_1 y + C_1 z + D_1 = 0 \tag{3.5a}$$

$$A_2 x + B_2 y + C_2 z + D_2 = 0 \tag{3.5b}$$

One point of the given line can be found upon specifying one of its three coordinates, the remaining two being found upon solving the system of equations (3.5a & 3.5b) for those coordinates.

3.1.3 Distance of a Point to a Plane

Given the plane Π represented by eq.(3.3b), we want to compute the distance of a *given* point $Q(\xi,\,\eta,\,\zeta)$ to the plane, as illustrated in Fig. 3.1. To do this, we proceed as in Subsection 2.2.1: We first find the unit normal $\boldsymbol{n}$ to the plane. Moreover, let $\boldsymbol{p}_0$, $\boldsymbol{q}$ and $\boldsymbol{p}$ be the position vectors of P_0, Q and $P(x,\,y,\,z)$, an arbitrary point of the plane. Since P_0 and P lie in the plane, the difference $\boldsymbol{p}-\boldsymbol{p}_0$ is perpendicular to the unit normal to the plane, $\boldsymbol{n}$, i.e.,

$$\boldsymbol{n}^T(\boldsymbol{p}-\boldsymbol{p}_0)=0 \tag{3.6}$$

or, in expanded form,

$$\boldsymbol{n}^T\boldsymbol{p}-\boldsymbol{n}^T\boldsymbol{p}_0=0 \tag{3.7}$$

Now, let us divide both sides of eq.(3.3b) by $\sqrt{A^2+B^2+C^2}$:

$$\frac{A}{\sqrt{A^2+B^2+C^2}}x+\frac{B}{\sqrt{A^2+B^2+C^2}}y+\frac{C}{\sqrt{A^2+B^2+C^2}}z+\frac{D}{\sqrt{A^2+B^2+C^2}}=0 \tag{3.8}$$

Comparison of eqs.(3.7) and (3.8) leads to

$$\boldsymbol{n}=\frac{1}{\sqrt{A^2+B^2+C^2}}\begin{bmatrix}A\\B\\C\end{bmatrix},\quad \boldsymbol{n}^T\boldsymbol{p}_0=-\frac{D}{\sqrt{A^2+B^2+C^2}} \tag{3.9}$$

From Fig. 3.1, the distance d sought is nothing but the *absolute value* of the projection of vector $\boldsymbol{q}-\boldsymbol{p_0}$ onto the unit normal $\boldsymbol{n}$, namely,

$$d=|\boldsymbol{n}^T(\boldsymbol{q}-\boldsymbol{p}_0)| \tag{3.10}$$

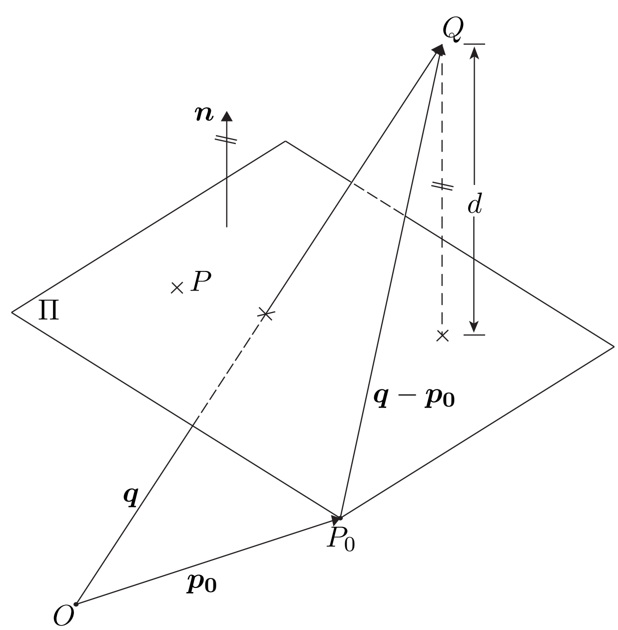


Figure 3.1: Distance of a point to a plane

3.1.4 Distance of a Point to a Line

A line $\mathcal{L}$ is given by the two planes (3.5a & 3.5b). We want to find the distance of a point $Q(\xi, \eta, \zeta)$ to $\mathcal{L}$. First, we need a unit vector $\boldsymbol{e}$ parallel to $\mathcal{L}$ and a point P_0 of $\mathcal{L}$. If we denote by $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ the unit normals to the two planes, then we can obtain $\boldsymbol{e}$ as $\boldsymbol{n}_1 \times \boldsymbol{n}_2 / \|\boldsymbol{n}_1 \times \boldsymbol{n}_2\|$. Moreover, $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ are produced using the expression for $\boldsymbol{n}$ displayed in eq.(3.9):

$$\boldsymbol{n}_i = \frac{1}{\sqrt{A_i^2 + B_i^2 + C_i^2}} \begin{bmatrix} A_i \\ B_i \\ C_i \end{bmatrix}, \quad i = 1, 2 \tag{3.11}$$

We thus have the layout of Fig. 3.2(a):

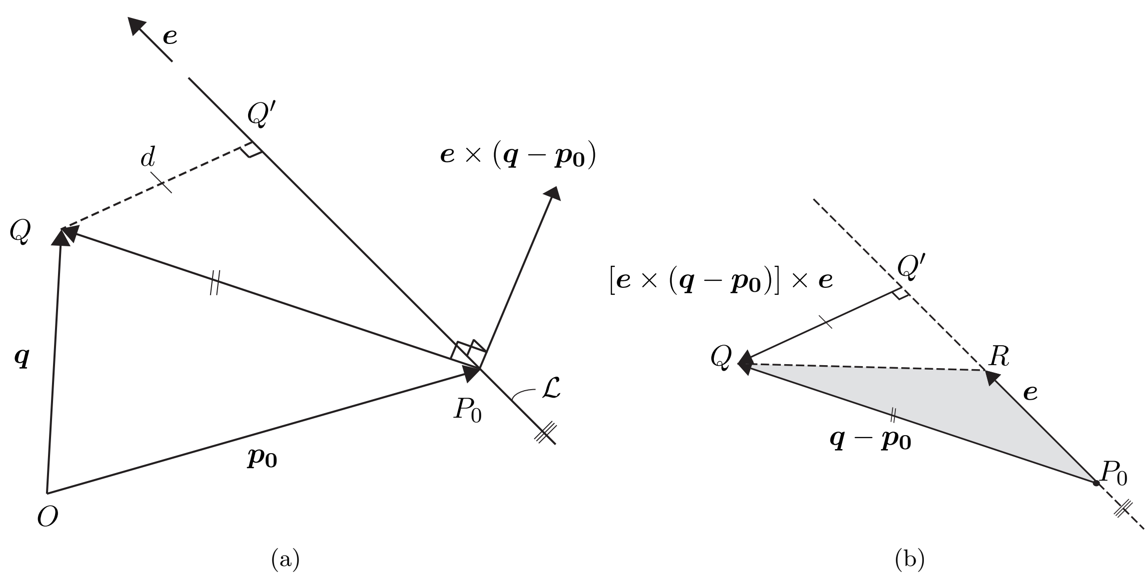


Figure 3.2: Distance of a point to a line: (a) general layout; (b) geometric interpretation of $[\boldsymbol{e} \times (\boldsymbol{q} - \boldsymbol{p_0})] \times \boldsymbol{e}$

From Fig. 3.2(a), P_0 is a point of $\mathcal{L}$, of position vector $\boldsymbol{p}_0$, while $\boldsymbol{e} \times (\boldsymbol{q} - \boldsymbol{p}_0)$ is a vector normal to the plane defined by $\mathcal{L}$ and Q, its norm $\|\boldsymbol{e} \times (\boldsymbol{q} - \boldsymbol{p}_0)\|$ being twice the area of the triangle P_0RQ depicted in Fig. 3.2(b). Moreover, if we regard P_0R as the base of the triangle, d becomes its height, and hence,

$$\overline{P_0R}d = \|\boldsymbol{e} \times (\boldsymbol{q} - \boldsymbol{p}_0)\| \tag{3.12}$$

where $\overline{P_0R} = \|\boldsymbol{e}\| = 1$, so that

$$d = \|\boldsymbol{e} \times (\boldsymbol{q} - \boldsymbol{p}_0)\| \tag{3.13}$$

thereby computing the desired length. Notice, moreover, that

$$\boldsymbol{q} - \boldsymbol{q}' = [\boldsymbol{e} \times (\boldsymbol{q} - \boldsymbol{p}_0)] \times \boldsymbol{e} \tag{3.14}$$

as the reader can readily verify, in which $\boldsymbol{q}'$ is the position vector of Q'.

3.1.5 Distance Between Two Skew Lines

We have the general layout of Fig. 3.3, depicting two *skew* lines $\mathcal{L}_1$ and $\mathcal{L}_2$, parallel to the unit vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ and passing through P_1 and P_2, respectively. If $\boldsymbol{n}$ denotes the unit normal to $\mathcal{L}_1$ and $\mathcal{L}_2$, then, apparently, the distance d between the two lines is nothing but the absolute value of the projection of $\boldsymbol{p}_2 - \boldsymbol{p}_1$ onto $\boldsymbol{n}$, i.e.,

$$d = |\boldsymbol{n}^T(\boldsymbol{p}_2 - \boldsymbol{p}_1)|, \quad \boldsymbol{n} \equiv \frac{\boldsymbol{e}_1 \times \boldsymbol{e}_2}{||\boldsymbol{e}_1 \times \boldsymbol{e}_2||} \tag{3.15}$$

The above relations are illustrated in Fig. 3.3(b)

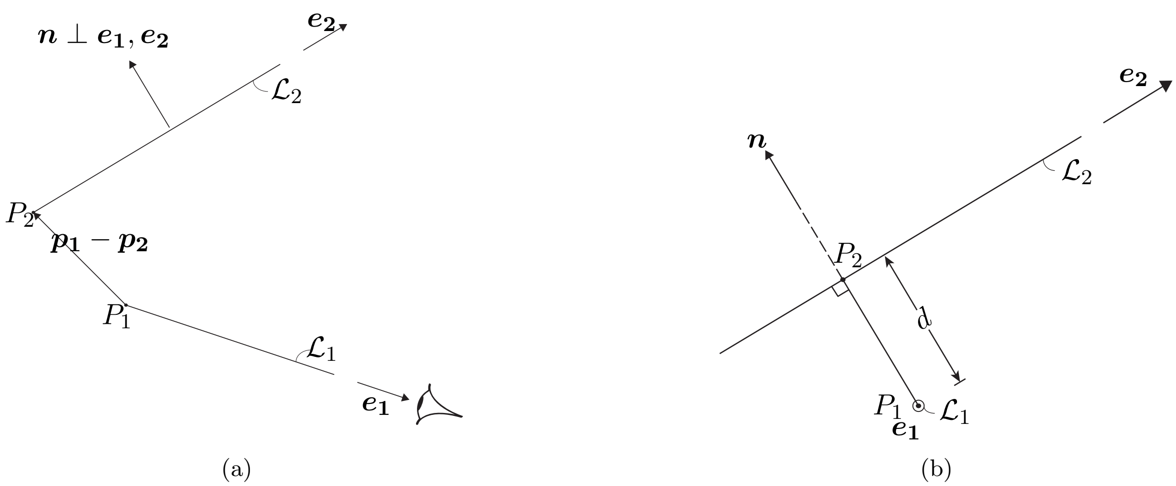


Figure 3.3: Distance between two lines: (a) general layout; (b) view with $\mathcal{L}_1$ projected as a point

3.2 Surfaces

A surface is a two-dimensional set of points, extending in two directions that change from point to point, but has no thickness. We will study various types of surfaces, as described below.

The *plane*, introduced and defined in Subsection 3.1.1, is the simplest surface. That is, the plane is the *perpendicular bisector* of the *segment* defined by the two points. A plane can also be visualized as a set of lines passing through a given point and perpendicular to one given direction. In *computer graphics*, *solid objects* can be bounded by planes, forming *facets* of the solid, each facet being a polygon. In this case, the solid turns out to be a *polyhedron*. As a matter of fact, arbitrary surfaces, like airplane fuselages, are sometimes approximated, for certain computations pertaining to the solids that they enclose—volume, centroid-location, etc.—by polyhedra.

In increasing order of *complexity*, the next surface is the *quadric*, namely, a surface defined, in a certain coordinate frame, by a *quadratic tri-variate polynomial*, namely,

$$F(x,\, y,\, z) = A_{11}x^2 + 2A_{12}xy + 2A_{13}xz + A_{22}y^2 \tag{3.16}$$

$$+ 2A_{23}yz + A_{33}z^2 + B_1x + B_2y + B_3z + C = 0 \tag{3.17}$$

which can be cast in the compact form

$$F(\boldsymbol{p}) = \boldsymbol{p}^T\boldsymbol{A}\boldsymbol{p} + \boldsymbol{b}^T\boldsymbol{p} + C \tag{3.18a}$$

where

$$\boldsymbol{A} \equiv \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{12} & A_{22} & A_{23} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}, \quad \boldsymbol{b} \equiv \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} \tag{3.18b}$$

and $\boldsymbol{p}$ defined as in eq.(3.1). The above expression can be cast in array form, similar to eq.(2.12), if we introduce homogeneous coordinates, namely,

$$F(p) = \boldsymbol{p}^T\boldsymbol{R}\boldsymbol{p} = 0 \tag{3.19a}$$

where

$$\boldsymbol{p} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} matA & \boldsymbol{b} \\ vecb^T & C \end{bmatrix} \tag{3.19b}$$

Other quadrics are the *ellipsoid*, a particular case of which is the *sphere*; the *two-sheet hyperboloid*; the *single-sheet hyperboloid*; the *hyperbolic paraboloid*; and the *paraboloid*. In the same way that we have criteria based on the entries of the upper-left block of matrix $\boldsymbol{R}$ of eq.(2.13) to characterize the conics, there are criteria to characterize the quadric at hand. These criteria are based on matrix $\boldsymbol{A}$ of eq.(3.19b). However, these criteria fall outside the scope of this course, and will not be pursued here.

Out of the foregoing surfaces, the single-sheet hyperboloid deserves special attention, as it leads, as special cases, to well-known familiar surfaces such as the cylinder and the cone. A single-sheet hyperboloid can be generated by the motion of a line constrained to intersect three given *skew lines*. Two lines are said to be skew when they do not intersect, which means that they do not intersect at all, not even at infinity, which is the case of parallel lines. Shown in Fig. 3.4 is a picture of a single-sheet hyperboloid, as defined by three skew lines.

All surfaces derived from the motion of a line belong to the class of *ruled surfaces*. Ruled surfaces are generated by the motion of a line, termed the *generatrix*, along a curve termed the *directrix*, the relative orientation of the generatrix with respect to the directrix being, in general, variable. Depending on the pattern of this variation, different surfaces can be obtained from the same directrix. In particular, when the orientation of the generatrix is kept constant, a *cylindrical surface* is obtained. Obviously, if the directrix is a circle and the generatrix remains normal to the plane of the circle, then a *circular cylinder* is generated. If the directrix is still a circle, but the generatrix is constrained to pass through a given point,

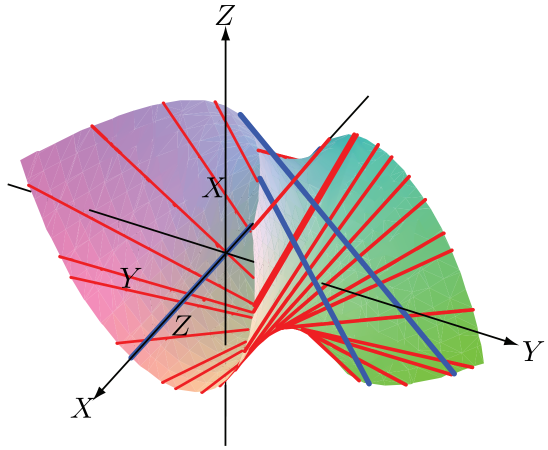


Figure 3.4: A single sheet hyperboloid created using three skew lines

then a cone is obtained. Moreover, if the given point lies on the normal to the plane of the circle passing through the centre of the circle, then a right circular cone is obtained.

Ruled surfaces are *curved* in one direction, that of the directrix, but are straight in the direction of the generatrix. For this reason, such surfaces are sometimes termed *single-curve surfaces*. The most general surfaces are *double-curve*. Examples of these are the ellipsoid and the paraboloid.

In computer graphics, certain special kinds of surfaces can be generated from a generatrix. For example, to generate an axially symmetric surface, like the sphere, we can turn a circle around one of its diameters. Such surfaces are termed *surfaces of revolution*. Other form of generating surfaces is by *extrusion*, whereby a generatrix is translated in one fixed direction.

A double-curve surface is sometimes referred to as a *warped surface*.

3.3 Simple Solids

3.3.1 Cones

Description

A conic surface is a ruled surface formed by a line (generatrix) moving along a curved path such that the line always passes through a fixed point, called the vertex. Each position of the generatrix is called an element of the surface. The faces of the teeth of a bevel gear are made of conic surfaces. The simplest conic surfaces are those whose generatrix makes a constant angle with a fixed line, called its *axis*. Such surfaces, when cut by a plane, form cones. We have three basic classes of cones, as illustrated in Fig. 3.5:

- If the axis is perpendicular to the base, the cone is called a *right cone*.
- If the axis is not perpendicular to the base, the cone is called *oblique*.
- If the cone is cut off, we obtain a *truncated cone* or a *frustum* of a cone.

There are many applications for cones in engineering design, including: the nose cone of rockets; transition pieces for heating; ventilation and air-conditioning systems; and conical

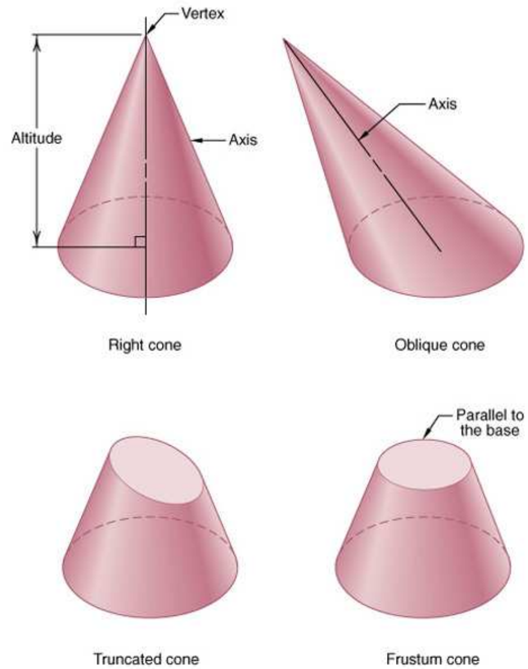


Figure 3.5: Classes of cones

roof sections. Cones are represented in multiview drawings by drawing the base curve, vertex, axis, and limiting elements in each view.

Moreover, a cone can also be defined as a bounded solid. For example, the representation of the right cone with apex at the origin, cone angle α and symmetric about the Y-axis ($y \geq 0$) can be derived from the geometry of Fig. 3.6.

$$x^2 + z^2 \leq k^2 y^2, \quad 0 \leq y \leq h, \quad k \equiv \tan\alpha \tag{3.20}$$

which we may write in array form as:

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -k^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \leq 0, \quad 0 \leq y \leq h \tag{3.21}$$

3.3.2 Cylinders

Description

A cylindrical surface is a ruled surface formed by a line segment called the *generatrix*, that moves while remaining parallel to a fixed line. Moreover, the generatrix moves so as to intersect a planar curve, called the *directrix*, contained in a plane intersecting the fixed line, called the axis, as shown in Fig. 3.7. The faces of the teeth of spur gears are cylindrical surfaces. When the distance between the generatrix and the axis is constant, we obtain a

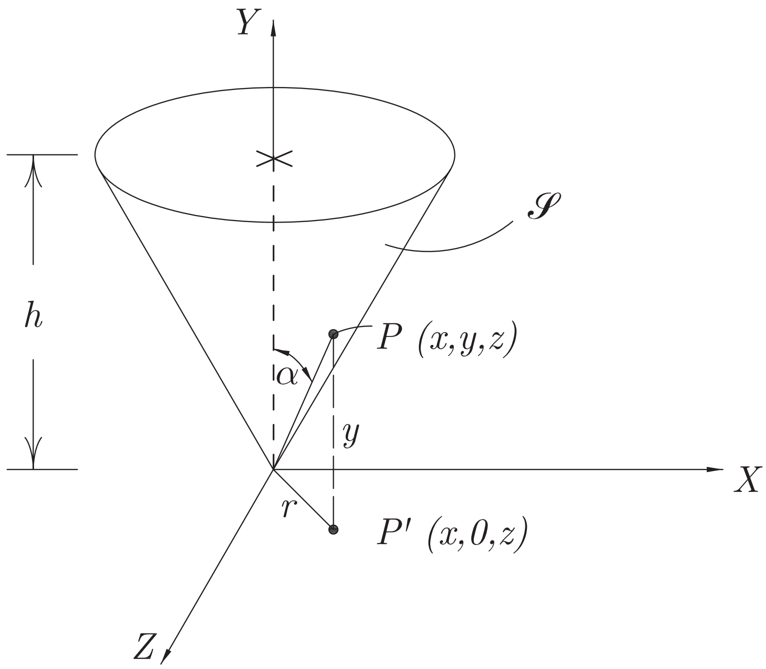


Figure 3.6: A solid, bounded cone

circular cylinder, or cylinder for brevity. If the axis is perpendicular to the base, the cylinder is *straight*; otherwise, the cylinder is *oblique*. A multiview drawing of a right circular cylinder shows the base curve (the directrix, which is a circle), the extreme elements, and the axis, as depicted in Fig. 3.7.

Moreover, as for the cone, a cylinder can be defined as a bounded body. For example, the equation of the cylinder whose axis is the Z-axis ($z \geq 0$) is defined by:

$$x^2 + y^2 \leq r^2, \quad 0 \leq z \leq h \tag{3.22}$$

which we may write in array form as:

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -r^2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \leq 0, \quad 0 \leq z \leq h \tag{3.23}$$

3.3.3 Regular Polyhedra

Regular polyhedra have regular polygons for faces. There are five regular polyhedra, also known as *Platonic solids*, namely, tetrahedron, hexahedron, octahedron, dodecahedron, and icosahedron. Illustrated in Fig. 3.8 are the Platonic solids.

Tetrahedron: A solid object with four equilateral triangular facets.

Hexahedron: A solid object with six quadrilateral facets.

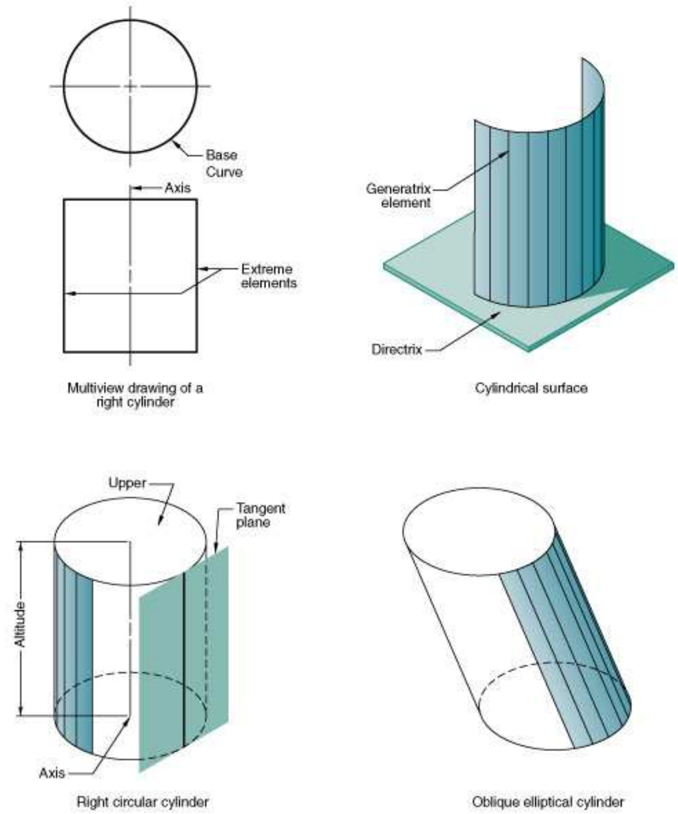


Figure 3.7: A sample of cylinders

Octahedron: A solid object with eight equilateral triangular facets.

Dodecahedron: A solid object with 12 regular pentagonal facets.

Icosahedron: A solid object with 20 equilateral triangular facets.

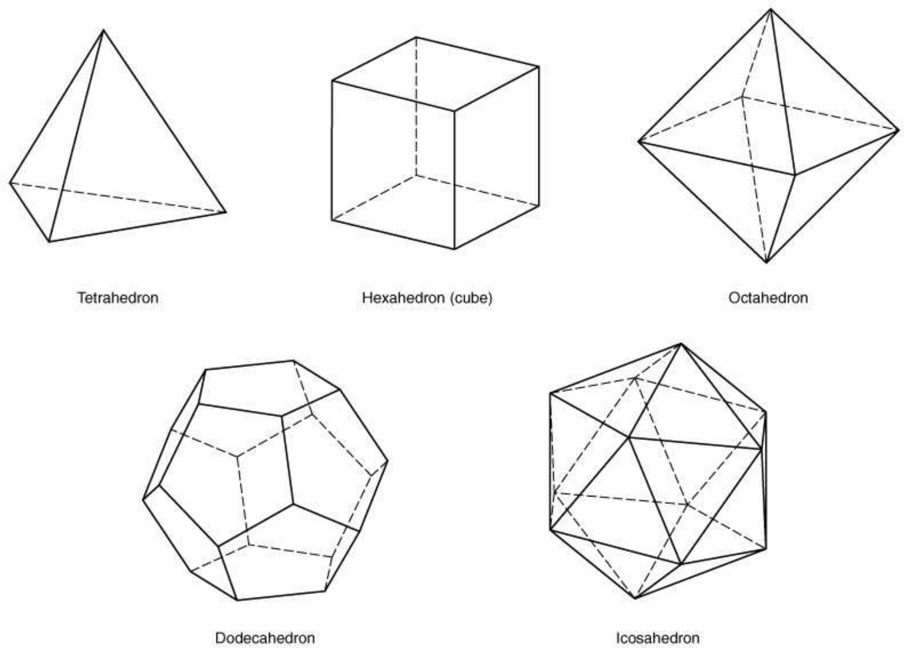


Figure 3.8: The Platonic solids (regular polyhedra)

3.3.4 Prisms and Pyramids

Prism: Description

A polygonal prism is a polyhedron that has two equal parallel faces, called its bases, and lateral faces that are parallelograms. The parallel bases are closed polygons of any shape and are connected by parallelogram sides. A line connecting the centres[1] of the two bases is called the axis. If the axis is perpendicular to the bases, the prism is *right*. If the axis is not perpendicular to the bases, the prism is *oblique*. A *truncated prism* is a prism that has a base not parallel to the other base. A parallelepiped is a prism with a rectangle or parallelogram as a base. Polygonal prisms are readily produced with 3D CAD software by using extrusion techniques. A classification of prisms is shown in Fig. 3.9.

Pyramid: Description

A pyramid is a polyhedron that has a polygon for a base and lateral faces that have a common intersection point called a vertex.

[1] The centre of a regular polygon is obvious, that of an irregular polygon is less so; the centre of the latter is defined as the *centroid* or *geometric centre* of the polygon. A plate of a homogeneous material can be suspended from the ceiling with a string attached to the centroid of the plate. When the plate is given an arbitrary orientation, this orientation is preserved, for the mass of the plate is concentrated at its centroid.

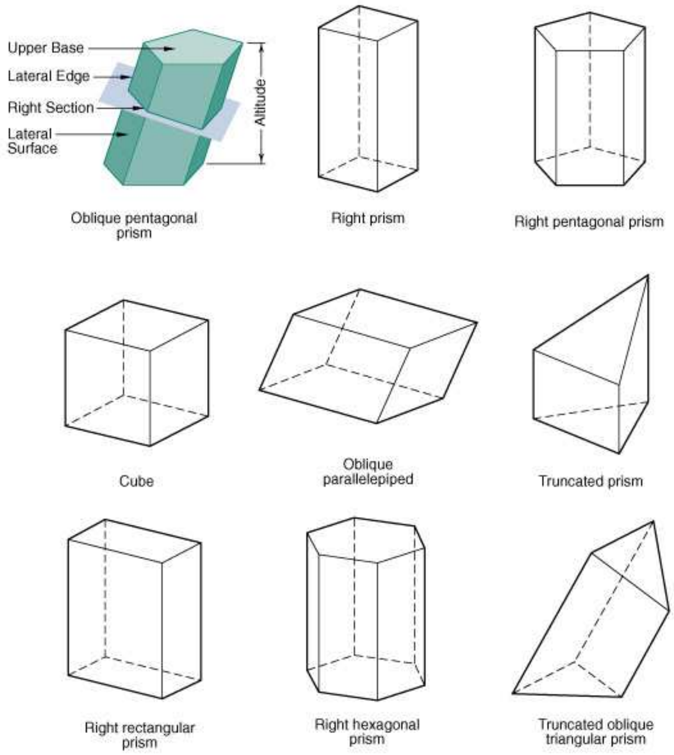


Figure 3.9: Classification of Prisms

3.4 Composite Solids: Boolean Operations

George Boole (1815–1864) invented the algebra we use for combining sets. The Boolean operators are: *union*, *intersection* and *difference*, as illustrated in Fig. 3.10.

The **union operator** $\cup$ combines two sets, $\mathcal{A}$ and $\mathcal{B}$ to form a third set $\mathcal{C}$ whose members are *either* members of $\mathcal{A}$ *or* members of $\mathcal{B}$. We express this as a Boolean algebraic relation:

$$\mathcal{C} = \mathcal{A} \cup \mathcal{B} \tag{3.24}$$

For example, if $\mathcal{A} = \{a, b, c, d\}$ and if $\mathcal{B} = \{c, d, e\}$, then $\mathcal{C} = \{a, b, c, d, e\}$.

The **intersection operator** $\cap$ combines two sets $\mathcal{A}$ and $\mathcal{B}$ to form a third set $\mathcal{C}$, whose members are members of *both* $\mathcal{A}$ *and* $\mathcal{B}$, which we write as

$$\mathcal{C} = \mathcal{A} \cap \mathcal{B} \tag{3.25}$$

Using the example of sets $\mathcal{A}$ and $\mathcal{B}$, whose members were described for the union operator, we find that $\mathcal{C} = \mathcal{A} \cap \mathcal{B} = \{c, d\}$. Apparently, the intersection of two sets that do not contain any common elements is *empty*. The empty set is represented as $\emptyset$. Two such sets are termed *disjoint*.

The **difference operator** combines two sets $\mathcal{A}$ and $\mathcal{B}$ to form a third set $\mathcal{C}$, whose members are only those of the first set that are not also members of the second. We write

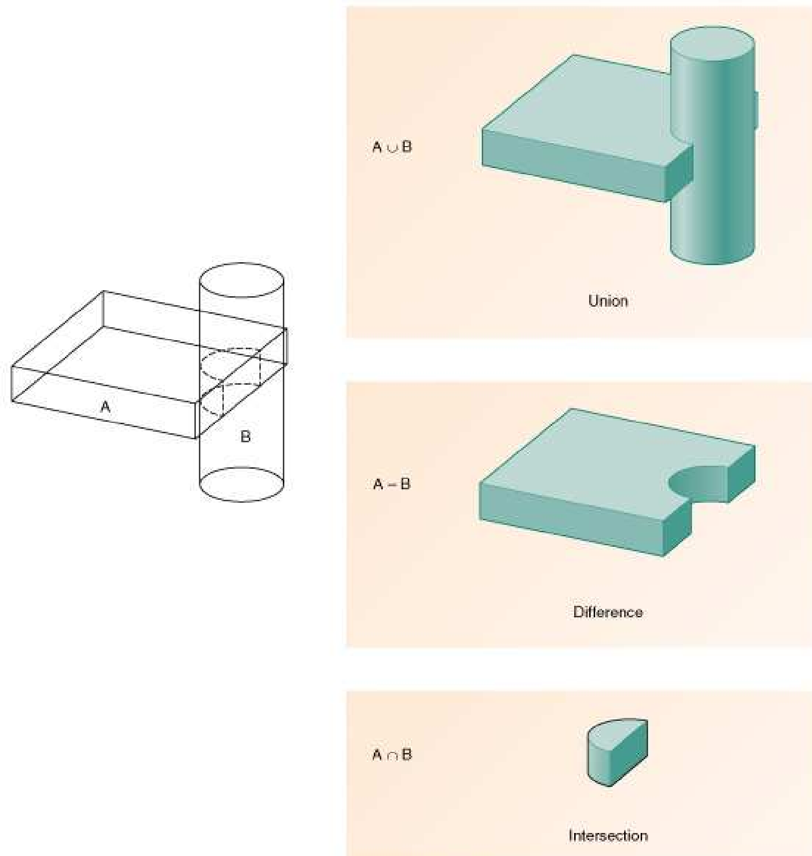


Figure 3.10: The three Boolean operations

this as

$$\mathcal{C} = \mathcal{A} - \mathcal{B} \tag{3.26}$$

Again, using the example of sets $\mathcal{A}$ and $\mathcal{B}$, whose members were described previously, we find that $\mathcal{C} = \mathcal{A} - \mathcal{B} = \{a, b\}$.
The Boolean operations can be used to adjoin primitives as shown in Fig. 3.11.

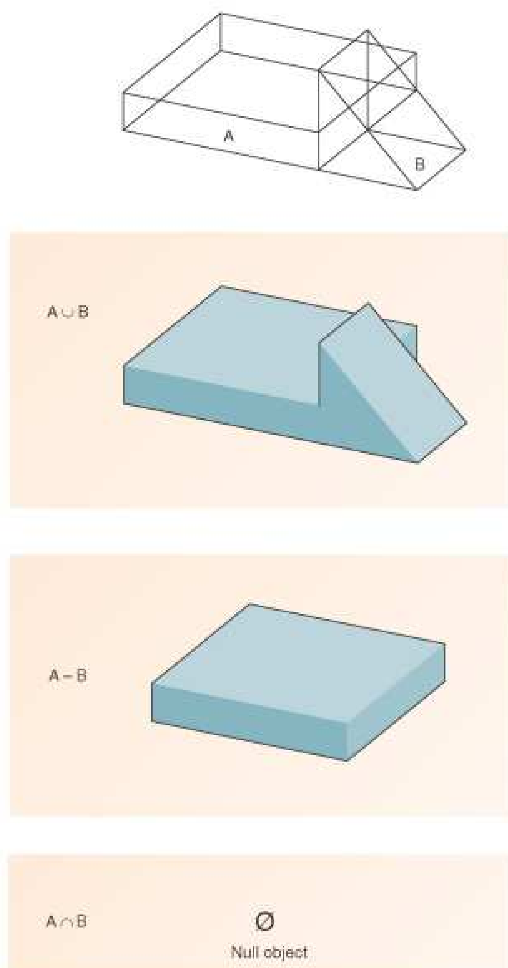


Figure 3.11: Boolean operations on adjoining primitives

The union and intersection operators are commutative, i.e.,

$$\mathcal{A} \cup \mathcal{B} = \mathcal{B} \cup \mathcal{A} \tag{3.27}$$

and

$$\mathcal{A} \cap \mathcal{B} = \mathcal{B} \cap \mathcal{A} \tag{3.28}$$

As illustrated in Fig. 3.12, the difference operator is *not* commutative, for $\mathcal{A} - \mathcal{B} = \{a, b\}$ and $\mathcal{B} - \mathcal{A} = \{e\}$. Thus,

$$\mathcal{A} - \mathcal{B} \neq \mathcal{B} - \mathcal{A} \tag{3.29}$$

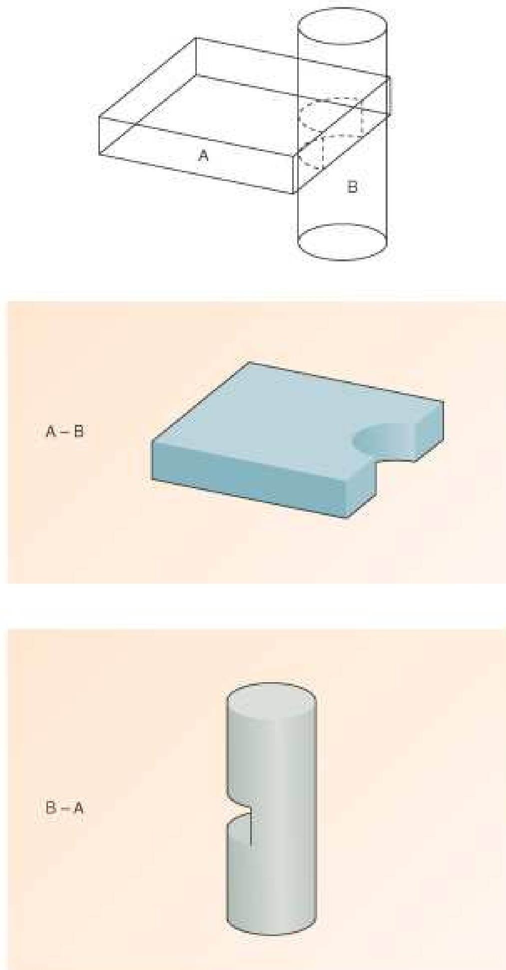


Figure 3.12: Effects of ordering of operands in a difference operation

Chapter 4

Affine Transformations

4.1 2D Transformations

Let $\boldsymbol{p}$ and $\boldsymbol{p}'$ denote the three-dimensional arrays containing the homogeneous coordinates of points P and P', respectively, in the XY plane. An affine transformation of P into P' is given by

$$\boldsymbol{p}' = \boldsymbol{T}\boldsymbol{p} \tag{4.1a}$$

with $\boldsymbol{p}$, $\boldsymbol{p}'$, and $\boldsymbol{T}$ given, in turn, by,

$$\boldsymbol{p} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad \boldsymbol{p}' = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}, \quad \boldsymbol{T} = \begin{bmatrix} \boldsymbol{M} & \boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{bmatrix} \tag{4.1b}$$

where $\boldsymbol{t}$ is the *translation vector*, $\boldsymbol{M}$ is a 2×2 matrix defining the type of transformation at hand and $\boldsymbol{0}$ is the 2-dimensional zero vector.

Notice an important property of affine transformations:

> *An affine transformation preserves parallelism, i.e., under an affine transformation applied to a figure, its parallel lines remain parallel after the transformation.*

4.1.1 Scaling

A *scaling transformation* allows an object to change by expanding or contracting its dimensions. Scaling constants in the x and y directions provide changes in length. If larger than unity, these constants represent expansion; if smaller than unity, they represent contraction. Scaling constants are always positive.

The scaling transformation of a point $P(x, y)$ into $P'(x', y')$ can be written as

$$\begin{aligned} x' &= S_x x \\ y' &= S_y y \end{aligned} \tag{4.2}$$

In this case, then, matrix $\boldsymbol{M}$ and vector $\boldsymbol{t}$ of eq.(4.1b) become

$$\boldsymbol{M} = \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix}, \quad \boldsymbol{t} = \boldsymbol{0} \tag{4.3}$$

Scaling is said to be uniform, or *isotropic*, if the scaling factors in the x and y directions are equal. Figure 4.1 shows an example of uniform scaling, whereas Figure 4.2 is an example of nonuniform scaling, with a contraction in the horizontal direction.

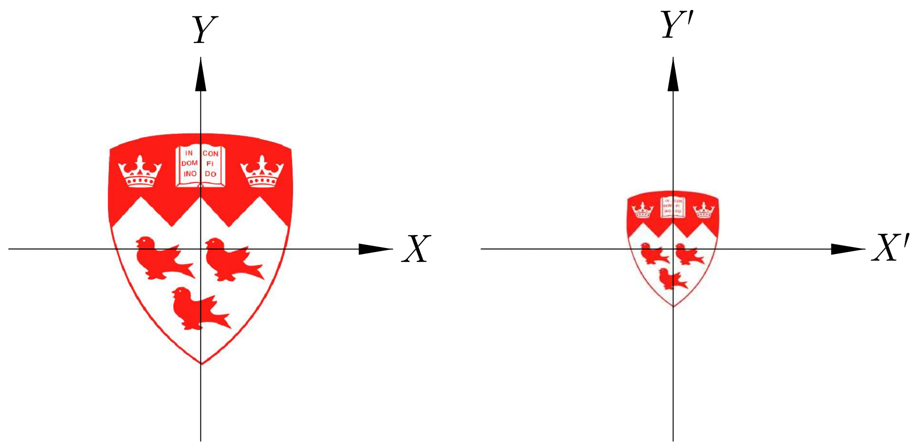


Figure 4.1: A uniform scaling

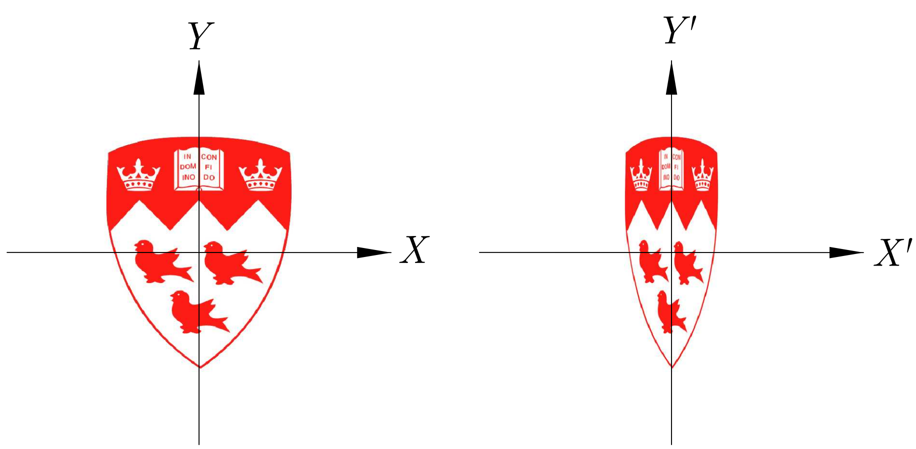


Figure 4.2: A nonuniform scaling

In some instances, scaling is needed about two orthogonal axes other than the given coordinate axes. In this case, if the axes intersect at the origin, then the affine transformation is obtained as a combination of scaling and rotation. If the two arbitrary orthogonal axes intersect at other point than the origin, then the rotation is accompanied by a translation. The most general affine transformation is studied in Subsection 4.1.5.

4.1.2 Translation

The ability to move parts of a model is an essential feature of any graphics system.

Translations cause an object to be displaced in a specific direction by a specific amount, while preserving its shape, size and orientation. The translation of the point $P(x, y)$ into

$P'(x', y')$ can be expressed as

$$x' = x + t_x$$
$$y' = y + t_y \tag{4.4}$$

In this case, matrix $\boldsymbol{M}$ and vector $\boldsymbol{t}$ of eq.(4.1b) become

$$\boldsymbol{M} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{t} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{4.5}$$

in which $\boldsymbol{M}$ is the 2×2 identity matrix because of shape-, size- and orientation-preservation.

The advantage of using homogeneous coordinates is apparent here: with Cartesian coordinates, rigid body translations could not be represented in *homogenous* form.

Figure 4.3 shows an example of translation.

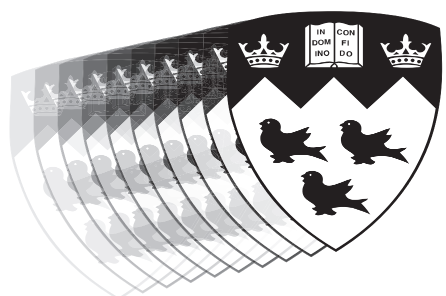


Figure 4.3: The McGill logo undergoing a translation.

4.1.3 Rotation

Rotation is frequently used to enable the viewer to see an object from different directions, or to assemble various geometric objects.

The rotation can be assumed to take place about the origin of the coordinate system by a specified angle θ. Should a rotation take place about a point other than the origin, then the corresponding transformation could be represented as a combination of translation and rotation.

Since we need a convention about the direction of rotation, we consider that counterclockwise rotations are positive, while their clockwise counterparts are negative.

We derive the rotation transformation via the polar coordinates of P, which are:

$$x = r\cos\phi$$
$$y = r\sin\phi \tag{4.6}$$

where ϕ is the angle and r the radius.

The transformed position P' of point P due to the rotation can be calculated by the use of simple trigonometric relations:

$$x' = r\cos(\phi + \theta) = r\cos\phi\cos\theta - r\sin\phi\sin\theta$$
$$y' = r\sin(\phi + \theta) = r\sin\phi\cos\theta + r\cos\phi\sin\theta \tag{4.7}$$

where x and y, as given by eq. (4.6), can be readily identified. Hence,

$$x' = x\cos\theta - y\sin\theta$$
$$y' = x\sin\theta + y\cos\theta \tag{4.8}$$

or, in array form,

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4.9}$$

Therefore, matrix $\boldsymbol{M}$ and vector $\boldsymbol{t}$ of eq.(4.1b) become

$$\boldsymbol{M} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad \boldsymbol{t} = \boldsymbol{0} \tag{4.10}$$

Figure 4.4 shows a rotation through an angle of $\theta = 45°$.

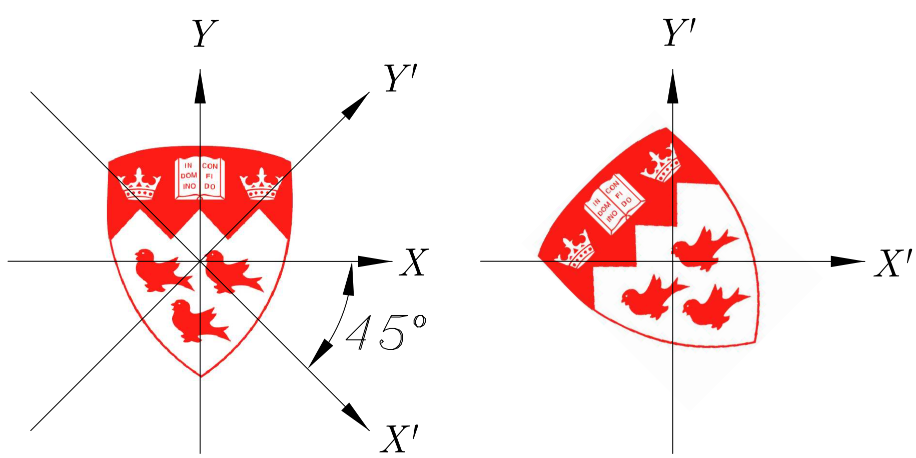


Figure 4.4: A rotation by $\theta = 45°$

4.1.4 Reflection

The concept of reflection can be understood by thinking of images in a mirror. The reflection transformation is useful in the construction of symmetric objects. For example, one half of a symmetric object may be created and then conveniently reflected to generate the whole object.

In 2D, reflections are defined about a line. The reflection matrix relative to either the X- or the Y-axes can be expressed in the form of eq.(4.1b), with $\boldsymbol{M}$ *improper orthogonal* and $\boldsymbol{t} = \boldsymbol{0}$. Improper orthogonality means that $\boldsymbol{M}$, besides being orthogonal, has a negative determinant, i.e.,

$$\boldsymbol{M}^T\boldsymbol{M} = \boldsymbol{M}\boldsymbol{M}^T = \boldsymbol{1}, \quad \det(\boldsymbol{M}) = -1 \tag{4.11}$$

Below we show different instances of reflections:

- About the X-axis,

$$\boldsymbol{M}_X = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \boldsymbol{t} = \boldsymbol{0} \tag{4.12}$$

This reflection is illustrated in Fig. 4.5.

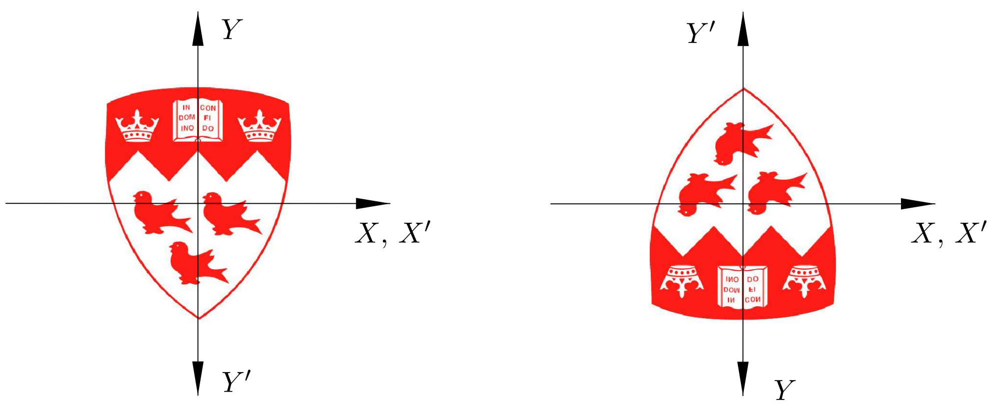


Figure 4.5: A reflection about the X-axis

- About the Y-axis,

$$\boldsymbol{M}_Y = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{t} = \boldsymbol{0} \tag{4.13}$$

This reflection is illustrated in Fig. 4.6.

- A composite reflection,

As the reader might expect, the combination of the two foregoing reflections is represented by the matrix

$$\boldsymbol{M} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \boldsymbol{t} = \boldsymbol{0} \tag{4.14}$$

which turns out to be a rotation about the origin through 180°, as depicted in Fig. 4.7.

Other reflections through arbitrary lines are also possible. For example, the reflection about the line $y = x$ is represented by:

$$\boldsymbol{M} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{t} = \boldsymbol{0} \tag{4.15}$$

An example of reflection about the line $y = x$ is illustrated in Fig. 4.8.

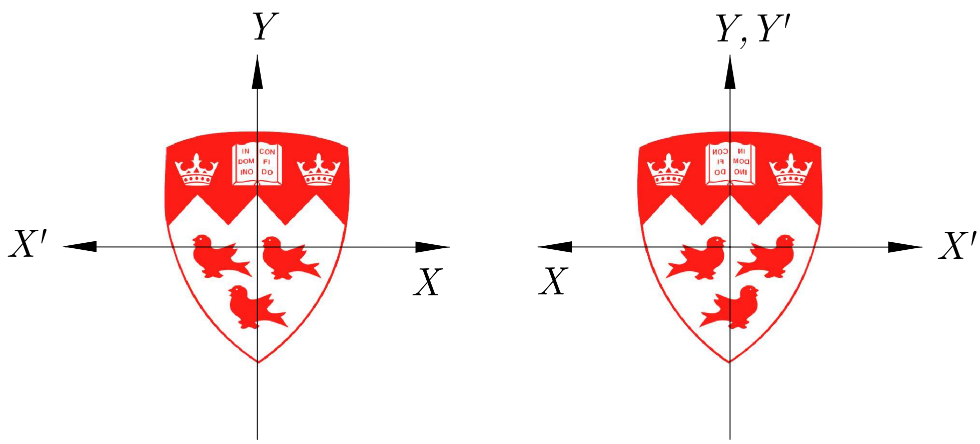


Figure 4.6: A reflection about the Y-axis

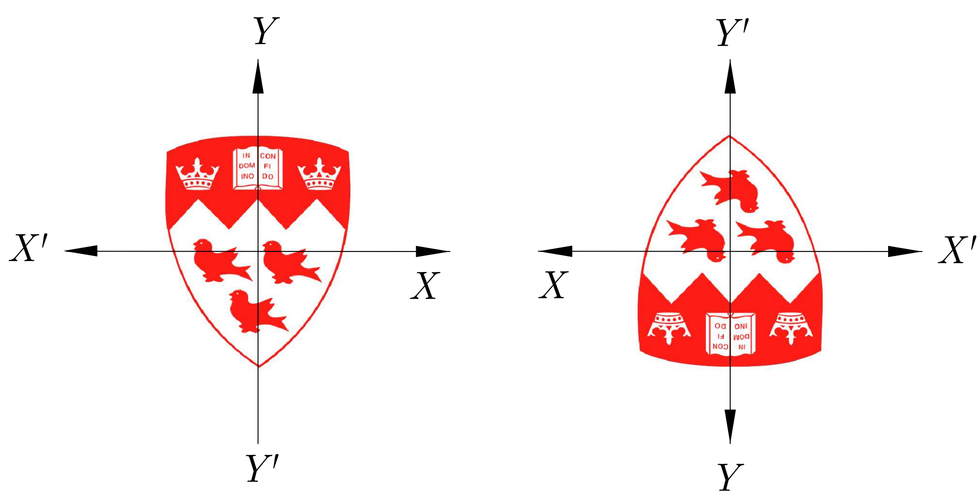


Figure 4.7: The composition of one reflection about the X-axis with one about the Y-axis, equivalent to a rotation about the origin through 180°

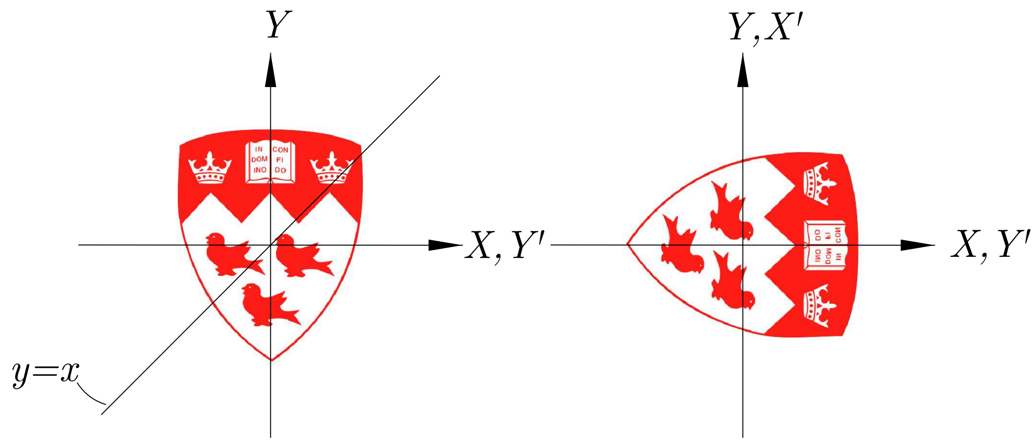


Figure 4.8: A reflection about the line $y = x$

4.1.5 Scaling Along Two Arbitrary Orthogonal Axes

We shall derive in this subsection the expression for the 3×3 matrix of the homogeneous transformation that represents a nonuniform scaling along two arbitrary orthogonal axes. We consider the case in which the two axes intersect at the origin of the X-Y plane. Should the axes intersect at a point O' other than the origin, then the *equivalent homogeneous transformation* is obtained as a concatenation of a scaling along two arbitrary axes that intersect at the origin and a displacement from the origin to O'.

Let us label the arbitrary axes X' and Y', X' making an angle θ with the X-axis, in the clockwise direction. For brevity, we will denote with $\mathcal{F}$ the $\{O, X, Y\}$ frame, with $\mathcal{F}'$ its $\{O', X', Y'\}$ counterpart.

Without loss of generality, the desired expression will be derived by means of a deformation of the unit circle $\mathcal{C}$ centred at the origin of the two frames into an ellipse $\mathcal{E}$ of semiaxes $S_{x'}$ and $S_{y'}$, which are the scalings along the X'- and the Y'-axes, respectively. Further, if we now mimic the scaling matrix along the X- and Y-axes, as given by eq.(4.3), the homogeneous scaling matrix $\boldsymbol{T}_{s'}$ along the primed axes is given by

$$\boldsymbol{T}_{s'} = \begin{bmatrix} S_{x'} & 0 & 0 \\ 0 & S_{y'} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.16}$$

Moreover, the homogeneous transformation $\boldsymbol{T}_r$ that rotates the X-, Y-axes into their X', Y' counterparts is given by

$$\boldsymbol{T}_r = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.17}$$

As illustrated in Fig. 4.9, a point Q of $\mathcal{C}$ is transformed into a point R of $\mathcal{E}$ under the nonuniform scaling $\boldsymbol{T}_{s'}$. Let $\boldsymbol{q}$ and $\boldsymbol{r}$ denote, correspondingly, the three-dimensional arrays of homogeneous coordinates of points Q and R, i.e.,

$$\boldsymbol{q} = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}, \quad \boldsymbol{r} = \begin{bmatrix} S_{x'}x' \\ S_{y'}y' \\ 1 \end{bmatrix} \tag{4.18}$$

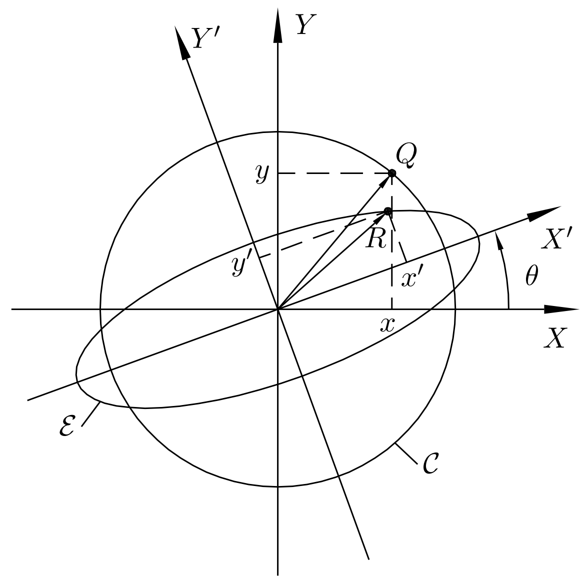


Figure 4.9: The nonuniform scaling of the unit circle centred at the origin along two orthogonal axes passing through the origin

However, $\boldsymbol{q}$ and $\boldsymbol{r}$ are arrays of frame-$\mathcal{F}'$ homogeneous coordinates, while the working frame is $\mathcal{F}$. Hence, a coordinate transformation is needed to go back to $\mathcal{F}$. This is readily done if we let $\boldsymbol{p}$ and $\boldsymbol{s}$ denote the three-dimensional arrays of homogeneous coordinates of Q and R in $\mathcal{F}$, respectively. Hence,

$$\boldsymbol{p} = \boldsymbol{T}_r\boldsymbol{q}, \quad \boldsymbol{s} = \boldsymbol{T}_r\boldsymbol{r} = \boldsymbol{T}_r\boldsymbol{T}_{s'}\boldsymbol{q} \tag{4.19a}$$

Further, from the first of the two foregoing equations, we can readily obtain

$$\boldsymbol{q} = \boldsymbol{T}_r^{-1}\boldsymbol{p} \tag{4.19b}$$

Upon substitution of eq.(4.19b) into the second equation of eq.(4.19a), we obtain

$$\boldsymbol{s} = \boldsymbol{T}_r\boldsymbol{T}_{s'}\boldsymbol{T}_r^{-1}\boldsymbol{p} \equiv \boldsymbol{T}_s\boldsymbol{p} \tag{4.19c}$$

where we have defined $\boldsymbol{T}_s$ as the homogeneous scaling transformation in the $\mathcal{F}$ frame, namely,

$$\boldsymbol{T}_s \equiv \boldsymbol{T}_r\boldsymbol{T}_{s'}\boldsymbol{T}_r^{-1} \tag{4.19d}$$

which is the affine transformation sought[1].

Of course, if the arbitrary orthogonal axes along which the scaling takes place do not intersect at the origin, but at a point of position vector $\boldsymbol{t}$, then the resulting affine transformation is obtained by *concatenating* the foregoing transformation with a pure translation $\boldsymbol{T}_t$, namely,

$$\boldsymbol{T}_t = \begin{bmatrix} \mathbf{1} & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{4.20}$$

[1]A relation of the form of eq.(4.19d) is known in Linear Algebra as a *similarity transformation*, which represents the change of entries of a matrix under a change of coordinate frame.

with $\mathbf{1}$ denoting the 2×2 identity matrix. The concatenated transformation then yields the equivalent transformation matrix

$$\boldsymbol{T}_{\rm tot} = \boldsymbol{T}_t \boldsymbol{T}_{\rm eq} \tag{4.21}$$

The foregoing transformation $\boldsymbol{T}_{\rm tot}$ is one of the two[2] most general affine transformations in 2D.

4.1.6 Examples

Example 4.1.1 Find the affine transformation that carries the unit circle centred at the origin of the XY-plane into the ellipse of semiaxes 1 and 2, centred at the point $C(3,\,2)$, with its major axis making an angle of 60° with the X-axis, as illustrated in Fig. 4.10. Finally, find the inverse transformation that carries the offset ellipse back into the unit circle centred at the origin of the X-Y plane.

Solution: We have a scaling about two orthogonal axes intersecting at a point offset from the origin. The scaling $\boldsymbol{T}_{s'}$ along these axes is given by

$$\boldsymbol{T}_{s'} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

while the rotation matrix $\boldsymbol{T}_r$ carrying the X-, Y-axes into axes X', Y', with X' making an angle of 60° with the X-axis, is given by

$$\boldsymbol{T}_r = \begin{bmatrix} 1/2 & -\sqrt{3}/2 & 0 \\ \sqrt{3}/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

whose inverse is readily found as

$$\boldsymbol{T}_r^{-1} = \begin{bmatrix} 1/2 & \sqrt{3}/2 & 0 \\ -\sqrt{3}/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Therefore, the scaling matrix in the X, Y frame is given by

$$\boldsymbol{T}_s = \boldsymbol{T}_r \boldsymbol{T}_{s'} \boldsymbol{T}_r^{-1} = \begin{bmatrix} 5/4 & \sqrt{3}/4 & 0 \\ \sqrt{3}/4 & 7/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

thereby obtaining the transformation carrying the unit circle centred at the origin into an ellipse of semiaxes 2 and 1 centred at the origin as well, and with its focal axis making an angle of 60° with the X-axis. To translate the foregoing ellipse so as to take its centre to point $C(3,\,2)$, we define below the homogeneous translation matrix $\boldsymbol{T}_t$:

$$\boldsymbol{T}_t = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

[2] The other equally general transformation is obtained if the frames $\mathcal{F}$ and $\mathcal{F}'$ are related by a reflection. We will not study this alternative affine transformation here.

The total homogeneous transformation is then given by

$$\boldsymbol{T}_{\text{total}} = \boldsymbol{T}_t \boldsymbol{T}_s = \begin{bmatrix} 5/4 & \sqrt{3}/4 & 3 \\ \sqrt{3}/4 & 7/4 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

The computation of the inverse homogeneous transformation is straightforward: Upon recalling eq. (1.91), we first have to compute $\boldsymbol{M}^{-1}$ and $-\boldsymbol{M}^{-1}\boldsymbol{t}$, the two upper blocks of $\boldsymbol{T}^{-1}$. These are readily obtained below, using the formulas for the inverse of 2×2 matrices:

$$\boldsymbol{M}^{-1} = \frac{1}{\Delta} \begin{bmatrix} 7/4 & -\sqrt{3}/4 \\ -\sqrt{3}/4 & 5/4 \end{bmatrix}, \quad \Delta = \frac{35}{16} - \frac{3}{16} = \frac{32}{16} = 2$$

whence

$$\boldsymbol{M}^{-1} = \begin{bmatrix} 7/8 & -\sqrt{3}/8 \\ -\sqrt{3}/8 & 5/8 \end{bmatrix}, \quad \boldsymbol{M}^{-1}\boldsymbol{t} = \begin{bmatrix} 7/8 & -\sqrt{3}/8 \\ -\sqrt{3}/8 & 5/8 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 21/8 - \sqrt{3}/4 \\ 5/4 - 3\sqrt{3}/8 \end{bmatrix}$$

Hence, the inverse homogeneous transformation sought is given by

$$\boldsymbol{T}^{-1} = \begin{bmatrix} 7/8 & -\sqrt{3}/8 & -21/8 + \sqrt{3}/4 \\ -\sqrt{3}/8 & 5/8 & -5/4 + 3\sqrt{3}/8 \\ 0 & 0 & 1 \end{bmatrix}$$

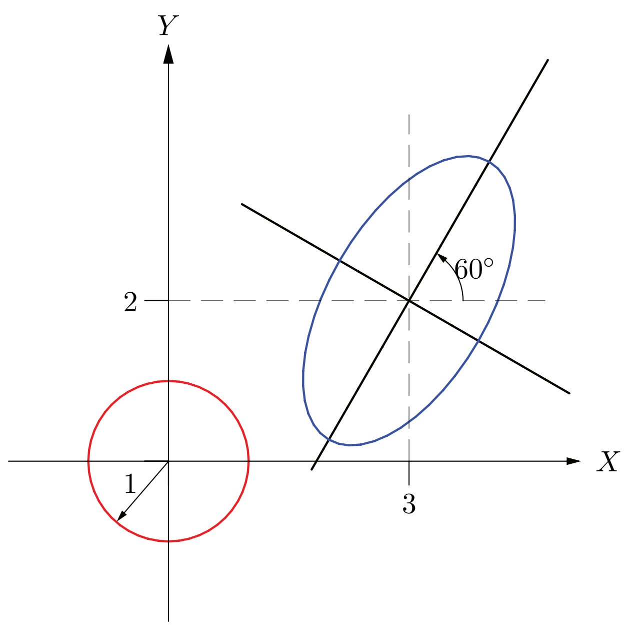


Figure 4.10: An affine transformation of the unit circle centred at the origin into an ellipse offset from the origin

Example 4.1.2 Given the cubic Lamé curve of Fig. 2.19(b), find the affine transformation required to squeeze it in such a way that the tangents at its inflection points (those at which the radius of curvature becomes infinite, namely the intersections of the curve with the coordinate axes) make an angle of $2\arccos(2\sqrt{5}/5) \approx 53.13°$. For quick reference, the original Lamé curve and its squeezed image are shown in Fig. 4.11.

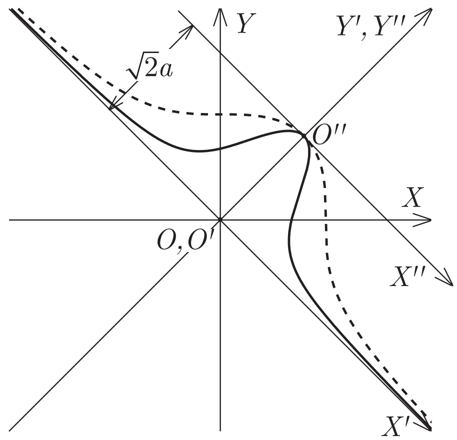


Figure 4.11: A squeezed cubic Lamé curve

Solution: It is not too difficult to show that, to transform the 90° angle made by the above-mentioned tangents of the original curve to $2\arccos(2\sqrt{5}/5)$, a scaling of 1/2 in a direction normal to the axis of symmetry is needed. Now, the axis of symmetry of the curve is a line passing through the origin and making an angle of 45° with the X-axis, its normal thus making an angle of $-45°$ with the same axis. Let the normal be labelled X', the axis of symmetry Y'.

The homogeneous transformation $\boldsymbol{T}_s$ representing the desired squeezing is given by the product

$$\boldsymbol{T}_s = \boldsymbol{T}_r \boldsymbol{T}_{s'} \boldsymbol{T}_r^{-1}$$

where $\boldsymbol{T}_{s'}$ is a nonuniform scaling along the X'- and Y'-axes, while $\boldsymbol{T}_r$ a cw rotation of 45°, namely,

$$\boldsymbol{T}_{s'} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{T}_r = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hence, the homogeneous transformation $\boldsymbol{T}_s$ in the X-, Y-axes becomes

$$\boldsymbol{T}_s = \boldsymbol{T}_r \boldsymbol{T}_{s'} \boldsymbol{T}_r^{-1} = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 3/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

thereby obtaining the desired scaling transformation.

Example 4.1.3 Find the affine transformation that a) carries the cubic Lamé curve, displayed in Fig. 2.19(b), into a configuration whereby: (a) its "hunch" lies at the origin of the X-Y plane; (b) its tangent at the "hunch" coincides with the X axis; and (c) its tangents at the inflection points make an angle of $2\cos^{-1}(\sqrt{5}/5) \approx 53.3°$, as displayed in Fig. 4.12.

Solution: With reference to Fig. 4.11, the curve is transformed by (a) a distortion under which the curve is squeezed in the X''-direction, while its dimensions in the Y''-direction are preserved, and (b) a displacement taking the $\{O'', X'', Y''\}$ frame to a configuration in which it coincides with the $\{O, X, Y\}$ frame.

The scaling matrix $\boldsymbol{T}_s$ that produces the squeezed Lamé curve was derived in Example 4.1.2. We reproduce this matrix below for quick reference:

$$\boldsymbol{T}_s = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 3/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The displacement is represented by a 3×3 homogeneous-transformation matrix $\boldsymbol{T}_d$ given by a rotation matrix through an angle of $+45°$ and a translation from O'' to O. We studied in Example 4.1.1 the displacement from frame $\{O, X, Y\}$ to frame with origin at $C(2, 3)$ and axes rotated 50° with respect to the original frame; we call this a *direct displacement*— As a matter of fact, the displacement in Example 4.1.1 involved only one rotation, and no translation; in this example, a rotation and a translation are involved. Now we need the *inverse displacement* instead. We proceed by calculating first the direct displacement $\boldsymbol{T}$, the inverse displacement being obtained by matrix-inversion. Now, the direct displacement involves a rotation through $-45°$ and a translation from $O(0, 0)$ to $O''(a, a)$, the translation being represented by $\boldsymbol{t} = [a, a, 1]^T$. Hence,

$$\boldsymbol{T} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 & a \\ -\sqrt{2}/2 & \sqrt{2}/2 & a \\ 0 & 0 & 1 \end{bmatrix}$$

The coordinates of O'' are calculated from the intersection of the cubic with line $y = x$. Substitution of the equation of this line into the cubic readily leads to

$$a = (1/2)^{1/3}$$

Hence,

$$\boldsymbol{T}_d = \boldsymbol{T}^{-1} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & -\sqrt{2}a \\ 0 & 0 & 1 \end{bmatrix}$$

Hence, the equivalent affine transformation $\boldsymbol{T}_{\text{eq}}$ is given by

$$\boldsymbol{T}_{\text{eq}} = \boldsymbol{T}_d \boldsymbol{T}_s = \begin{bmatrix} \sqrt{2}/4 & -\sqrt{2}/4 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & -\sqrt{2}a \\ 0 & 0 & 1 \end{bmatrix}$$

thereby obtaining the desired transformation.

4.2 Computer Implementation of 2D Affine Transformations

The effect of rotating an object while leaving the coordinate axes fixed is equivalent to the effect of rotating the axes in the opposite direction by the same amount while leaving the

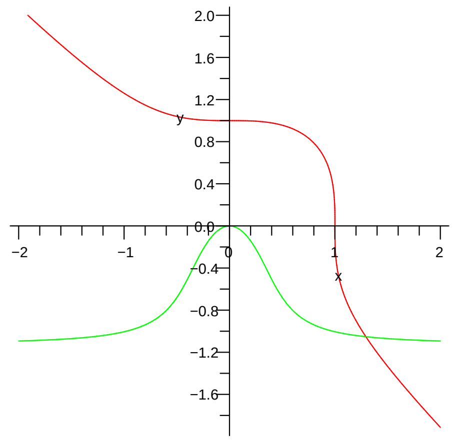


Figure 4.12: The affine transformation of a cubic Lamé curve into a displaced, squeezed configuration

object fixed. This observation is made apparent in Fig. 4.4: Rotating the McGill logo through $+45°$—i.e., ccw—while leaving the X-, Y-axes fixed is equivalent to rotating the axes through $-45°$—i.e., cw—while leaving the logo fixed.

As a matter of fact, the foregoing equivalence applies to *any* affine transformation. That is, when plotting the image of an object undergoing a given affine transformation, the position vector of any of its points has to be multiplied by the inverse of the given transformation.

4.2.1 Examples

Example 4.2.1 *Derive the affine transformation required to plot the ellipse of Example 4.1.1.*

Solution: To obtain the ellipse of Fig. 4.10, the vector $\boldsymbol{p}$ of homogeneous coordinates of an arbitrary point of the circle is multiplied by the inverse of $\boldsymbol{T}_{\text{total}}$, which was derived in Example 4.1.1, thereby producing a new vector $\boldsymbol{s}$ of homogeneous coordinates, namely,

$$\boldsymbol{s} = \boldsymbol{T}_{\text{total}}^{-1}\boldsymbol{p} = \begin{bmatrix} 7/8 & -sqrt3/8 & -21/8+\sqrt{3}/4 \\ -\sqrt{3}/8 & 5/8 & -5/4+3\sqrt{3}/8 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} (7/8)x-(\sqrt{3}/8)y-21/8+\sqrt{3}/4 \\ -(\sqrt{3}/8)x+(5/8)y-5/4+3\sqrt{3}/8 \\ 1 \end{bmatrix}$$

Hence, the equation of the displaced ellipse is obtained as

$$\begin{aligned} E &= q_1^2+q_2^2-1 \\ &= \left(\frac{7}{8}x-\frac{\sqrt{3}}{8}y-\frac{21}{8}+\frac{\sqrt{3}}{4}\right)^2+\left(-\frac{\sqrt{3}}{8}x+\frac{5}{8}y-\frac{5}{4}+\frac{3\sqrt{3}}{8}\right)^2-1 \\ &= \frac{13}{16}x^2-\frac{3\sqrt{3}}{8}xy+\frac{7}{16}y^2+\frac{1}{8}(6\sqrt{3}-39)x+\frac{1}{8}(9\sqrt{3}-14)y+\frac{1}{16}(129-36\sqrt{3})=0 \end{aligned}$$

Example 4.2.2 *Obtain the affine transformation required to plot the squeezed Lamé curve of Fig. 4.11.*

Solution: The squeezed cubic Lamé curve of Fig. 4.11 was obtained by multiplying the vector $\boldsymbol{p}$ of homogeneous coordinates of a point of the original Lamé curve by $\boldsymbol{T}_s^{-1}$, for $\boldsymbol{T}_s$ given as in Example 4.1.2, thereby obtaining the new vector $\boldsymbol{q}$, namely,

$$\boldsymbol{q} = \boldsymbol{T}_s^{-1}\boldsymbol{p} = \begin{bmatrix} 3/2 & -1/2 & 0 \\ -1/2 & 3/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} (3/2)x - (1/2)y \\ -(1/2)x + (3/2)y) \\ 1 \end{bmatrix}$$

Hence, the equation of the squeezed curve is

$$\begin{aligned} L_3 &= q_1^3 + q_2^3 - 1 \\ &= \frac{13}{4}x^3 - \frac{9}{4}x^2y - \frac{9}{4}xy^2 + \frac{13}{4}y^3 - 1 = 0 \end{aligned}$$

Example 4.2.3 *Find the affine transformation required for plotting the squeezed and displaced cubic Lamé curve of Example 4.1.3, and plotted in Fig. 4.12.*

Solution: We need the inverse of $\boldsymbol{T}_{\rm eq}$, which was computed in the examlpe cited above, to obtain a new vector $\boldsymbol{q}$ of homogeneous coordinates of a point of the displaced, squeezed curve from its original vector $\boldsymbol{p} = [\,x,\, y,\, 1\,]^T$:

$$\boldsymbol{q} = \boldsymbol{T}_{\rm eq}^{-1}\boldsymbol{p} = \begin{bmatrix} \sqrt{2} & \sqrt{2}/2 & a \\ -\sqrt{2} & \sqrt{2}/2 & a \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \sqrt{2}x + (\sqrt{2}/2)y + a \\ -\sqrt{2}x + (\sqrt{2}/2)y + a \\ 1 \end{bmatrix}$$

Hence, the cubic equation becomes

$$\begin{aligned} G &= \left(\sqrt{2}x + \frac{\sqrt{2}}{2}y + a\right)^3 + \left(-\sqrt{2}x + \frac{\sqrt{2}}{2}y + a\right)^3 - 1 \\ &= 6\sqrt{2}x^2y + 6(2^{2/3})x^2 + \frac{\sqrt{2}}{2}y^3 + \frac{3}{2}2^{2/3}y^2 + \frac{3}{2}2^{5/6}y = 0 \end{aligned}$$

where a was substituted by its numerical value $a = (1/2)^{3/2}$.

4.3 3D Transformations

Transformations in three-dimensional space are executed by the same methods used in two-dimensional space, with the addition of the z-coordinate. In homogeneous coordinates, these transformations are represented by a 4×4 matrix $\boldsymbol{T}$, mapping the 4-dimensional array $\boldsymbol{p}$ of homogeneous coordinates of P into its counterpart $\boldsymbol{p}'$ of P' in the form

$$\boldsymbol{p}' = \boldsymbol{T}\boldsymbol{p} \tag{4.22a}$$

where

$$\boldsymbol{T} = \begin{bmatrix} A & B & C & D \\ E & F & G & H \\ I & J & K & L \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{p} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad \boldsymbol{p}' = \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} \tag{4.22b}$$

Moreover, matrix $\boldsymbol{T}$ can be partitioned in the form:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{M} & \boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{bmatrix} \tag{4.23}$$

Matrix $\boldsymbol{M}$ in the upper left corner allows for scaling, reflection and rotation, while vector $\boldsymbol{t}$ accounts for translation. Lastly, $\boldsymbol{0}$ is the three-dimensional zero vector.

4.3.1 Scaling

The scaling transformation is obtained by placing values along the main diagonal of the general 4×4 transformation matrix. An arbitrary point $P(x, y, z, 1)$ is scaled to $P'(x', y', z', 1)$ by the transformation

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} A & 0 & 0 & 0 \\ 0 & F & 0 & 0 \\ 0 & 0 & K & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{4.24a}$$

In this case, matrix $\boldsymbol{M}$ is diagonal, while $\boldsymbol{t} = \boldsymbol{0}$, for there is no translation involved. Matrix $\boldsymbol{M}$ is, thus,

$$\boldsymbol{M} = \mathrm{diag}(A, F, K) \tag{4.24b}$$

This is an extension of two-dimensional scaling, described in Subsection 4.1.1. If the scaling factors A, F, K are not equal, the image of the object is distorted. Otherwise, a change in size occurs, but the original proportions are maintained.

In Fig. 4.13, a torus is uniformly scaled to form the smaller torus.

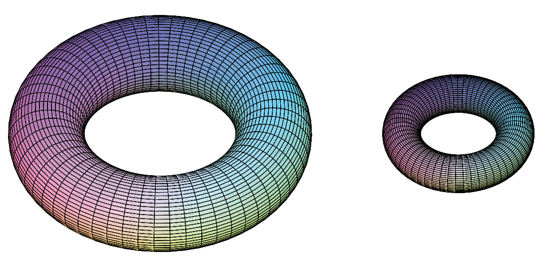


Figure 4.13: Uniform scaling

In Fig. 4.14, the torus on the left is scaled by (0.5, 1, 2) in the (x, y, z) directions, respectively, thus producing the surface of the right, which is no longer a surface of revolution. Notice that the central circle[3] of the torus becomes an ellipse of semi-axis of length ratio 0.5, while the cross sections become ellipses of variable semi-axis length ratios.

[3]This is the circle traced by the centre of the circle playing the role of the generatrix

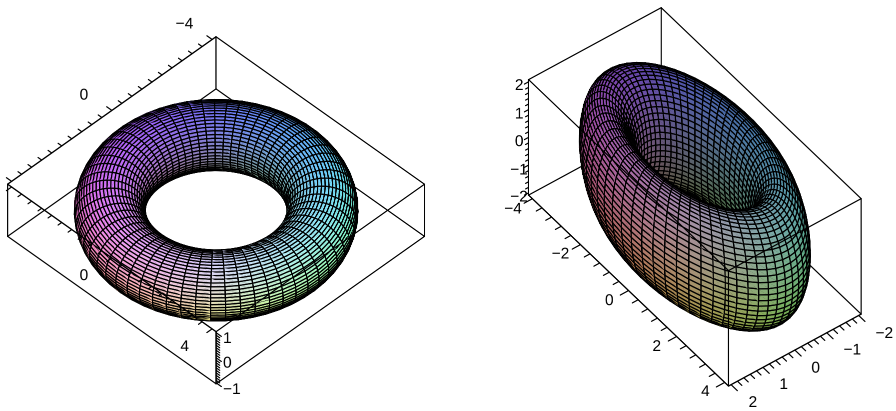


Figure 4.14: nonuniform scaling in 3D

4.3.2 Translation

A translation is a special case of rigid-body displacement, under which all the points of the body undergo the same displacement. The transformation translating a point $P(x, y, z, 1)$ to a new point $P'(x', y', z', 1)$ through (D, H, L) is given by:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & D \\ 0 & 1 & 0 & H \\ 0 & 0 & 1 & L \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{4.25a}$$

Notice that, in this case, all the points of the body undergo the same displacement, but the object is neither rotated nor distorted. We thus have

$$\boldsymbol{M} = \mathbf{1}, \quad \boldsymbol{t} = \begin{bmatrix} D & H & L \end{bmatrix}^T \tag{4.25b}$$

The values of D, H, L represent the relative translation of the point in the x, y, z directions, respectively.

In Fig. 4.15, we can see examples of translations.

4.3.3 Rotation

A rotation in 3D is another special case of rigid-body displacement. Under a rotation, the distance between every pair of object points is preserved and one point of the object remains stationary. The object is said to rotate about that point.

Rotations in three dimensions are more complex than their two-dimensional counterparts, because an axis of rotation, rather than a centre of rotation, must be specified. Rotations about an axis passing through the origin are characterized by a *proper orthogonal* matrix $\boldsymbol{M}$ and a zero translation, $\boldsymbol{t} = \mathbf{0}$. In this case, $\boldsymbol{M}$ has the properties below:

$$\boldsymbol{M}^T\boldsymbol{M} = \boldsymbol{M}\boldsymbol{M}^T = \mathbf{1}, \quad \det(\boldsymbol{M}) = +1 \tag{4.26}$$

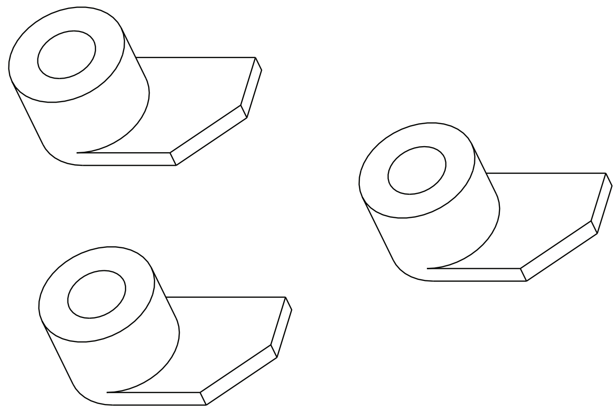


Figure 4.15: Translations in 3D

In particular, the rotation matrix for a Z-axis rotation through an angle θ is:

$$\boldsymbol{M}_Z = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.27}$$

which produces the mapping:

$$\begin{aligned} x' &= x\cos\theta - y\sin\theta \\ y' &= x\sin\theta + y\cos\theta \\ z' &= z \end{aligned} \tag{4.28}$$

In a similar manner, a rotation of θ about the Y-axis can be obtained by means of

$$\begin{aligned} x' &= x\cos\theta + z\sin\theta \\ y' &= y \\ z' &= -x\sin\theta + z\cos\theta \end{aligned} \tag{4.29}$$

and is correspondingly represented by

$$\boldsymbol{M}_Y = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \tag{4.30}$$

A rotation about the X-axis is:

$$\begin{aligned} x' &= x \\ y' &= y\cos\theta - z\sin\theta \\ z' &= y\sin\theta + z\cos\theta \end{aligned} \tag{4.31}$$

which is represented by

$$\boldsymbol{M}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{4.32}$$

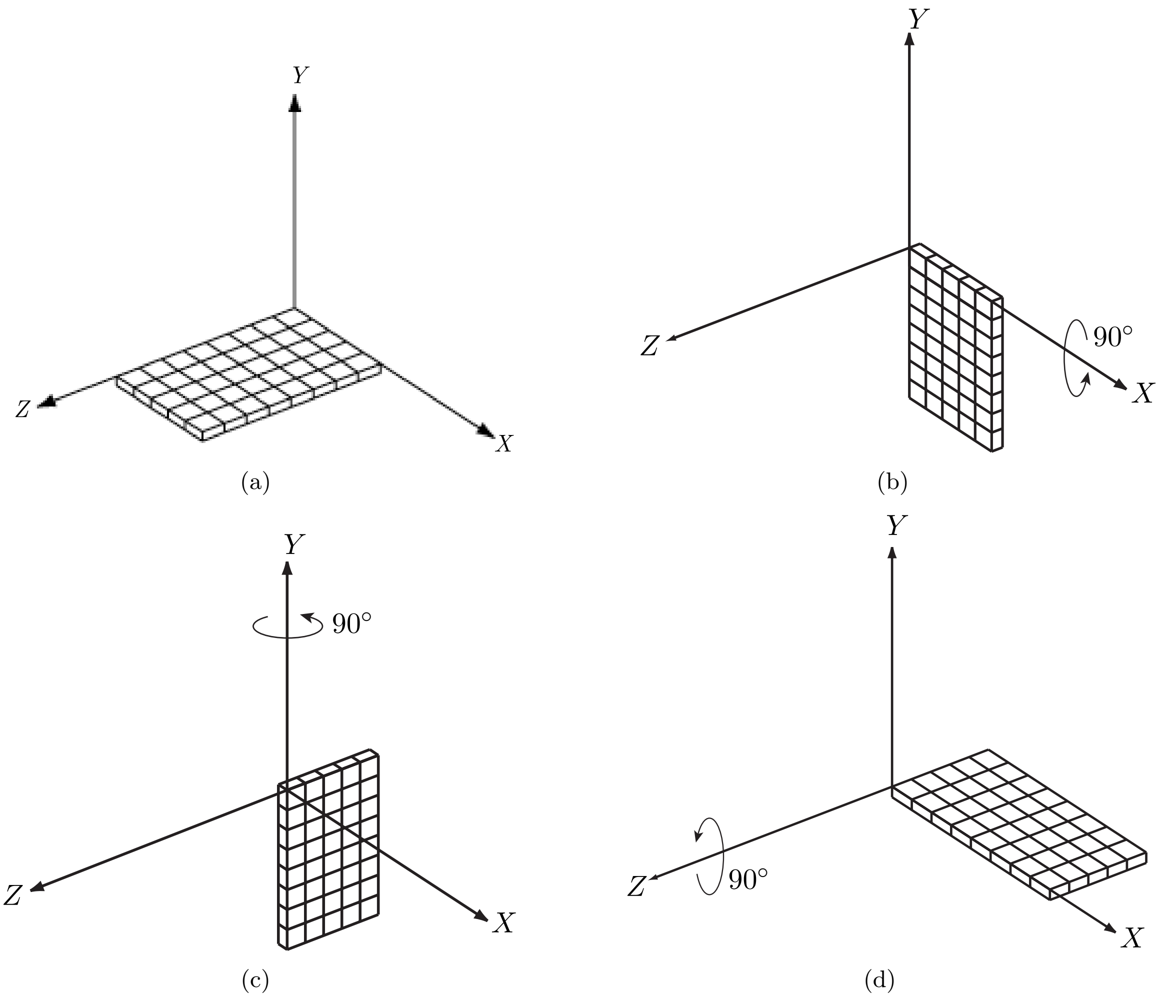


Figure 4.16: A solar panel: (a) in its original configuration; (b) after a rotation through 90° about the X-axis; (c) after a second rotation through 90° about the Y-axis; and (d) about a third rotation through 90° about the Z-axis

Sometimes, rotations about arbitrary axes are specified as a sequence of rotations about the coordinate axes, as illustrated with the solar panel of Fig. 4.16(a), used in telecommunications satellites to provide energy to their different instruments. In this case we have

$$\boldsymbol{M}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{M}_Y = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{M}_Z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

An arbitrary point P of coordinates (x, y, z) is carried into a new position P_1 of coordinates (x_1, y_1, z_1), after the first rotation. After the second rotation, P_1 is carried into P_2, of coordinates (x_2, y_2, z_2). The final position of P is P_3, of coordinates (x_3, y_3, z_3). Hence,

$$\boldsymbol{p}_1 = \boldsymbol{M}_X \boldsymbol{p}, \quad \boldsymbol{p}_2 = \boldsymbol{M}_Y \boldsymbol{p}_1, \quad \boldsymbol{p}_3 = \boldsymbol{M}_Z \boldsymbol{p}_2$$

whence

$$\boldsymbol{p}_3 = \boldsymbol{M}_Z \boldsymbol{M}_Y \boldsymbol{M}_X \boldsymbol{p} \equiv \boldsymbol{M}\boldsymbol{p}$$

That is, the total rotation $\boldsymbol{M}$ is obtained as

$$\begin{aligned}
\boldsymbol{M} &= \boldsymbol{M}_Z \boldsymbol{M}_Y \boldsymbol{M}_X \\
&= \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}
\end{aligned}$$

which, in this particular case, *happens* to be a rotation about Y through 90°.

The composition of rotations thus reduces to simple matrix multiplications. However, the inverse problem, given a rotation matrix, finding its axis and angle of rotation is less straightforward. We show how to accomplish this task with an example.

Example 4.3.1 *Matrix $\boldsymbol{M}$ shown below is claimed to represent a rotation of an object $\mathcal{B}$ rotating about the origin*

$$\boldsymbol{M} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

(a) Prove that the matrix indeed represents a rotation about the origin; then

(b) Find its axis and its angle of rotation

Solution:

(a) To represent a rotation about the origin, $\boldsymbol{M}$ must be proper orthogonal. We thus compute

$$\begin{aligned}
\boldsymbol{M}\boldsymbol{M}^T &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
\boldsymbol{M}^T\boldsymbol{M} &= \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}$$

whence $\boldsymbol{M}$ is indeed orthogonal. To be proper orthogonal, its determinant must be +1. We thus compute its determinant by expansion of cofactors of its first column:

$$\det(\boldsymbol{M}) = 0 + 0 + \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = +1$$

and hence, $\boldsymbol{M}$ indeed represents a rotation.

(b) As will become apparent presently, to find the axis of rotation of $\boldsymbol{M}$, we need the initial and final positions of two points of $\mathcal{B}$, none of which is the origin, for the origin does not move. Let us thus choose point A a distance d_A from the origin in the positive direction of X, and B a distance d_B in the positive direction of Y, as shown in Fig. 4.17.

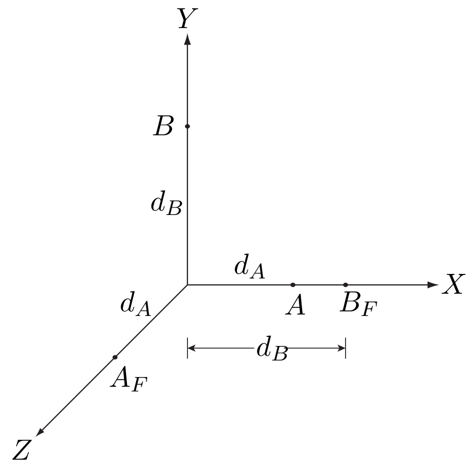


Figure 4.17: Two points, A and B, of an object $\mathcal{B}$ rotating about the origin, in the original and final configurations of $\mathcal{B}$, with the final point positions carrying the subscript F

Let, moreover, $\boldsymbol{a}$ and $\boldsymbol{b}$ be the position vectors of A and B, respectively, i.e.,

$$\boldsymbol{a} = \begin{bmatrix} d_A \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 0 \\ d_B \\ 0 \end{bmatrix}$$

In the final configuration of $\mathcal{B}$, A and B take the positions A_F and B_F, respectively of position vectors $\boldsymbol{a}_F$ and $\boldsymbol{b}_F$. That is,

$$\boldsymbol{a}_F = \boldsymbol{M}\boldsymbol{a} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} d_A \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ d_A \end{bmatrix}$$

$$\boldsymbol{b}_F = \boldsymbol{M}\boldsymbol{b} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ d_B \\ 0 \end{bmatrix} = \begin{bmatrix} d_B \\ 0 \\ 0 \end{bmatrix}$$

The axis of rotation is the set of points of $\mathcal{B}$ that do not change their position in the final configuration of the body. These points are known[4] to lie in a line $\mathcal{L}$ passing through the origin, as illustrated in Fig. 4.18

[4]Leonhard Euler proved this in 1776.

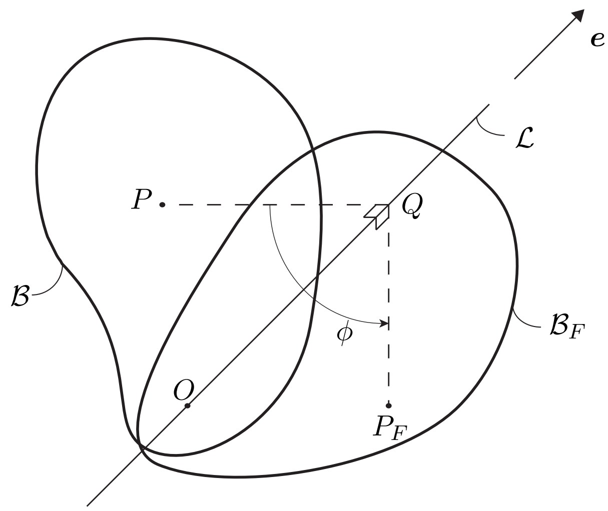


Figure 4.18: An object $\mathcal{B}$ in its original and final configurations; illustration of axis $\mathcal{L}$ and angle of rotation ϕ

Since a rotation entails no distortion, all points of $\mathcal{L}$ are equidistant from P and P_F. In particular, Q is the intersection of the perpendicular to $\mathcal{L}$ from P and P_F. Angle ϕ, measured as indicated in Fig. 4.18, is the angle of rotation. Notice that if vector $\boldsymbol{e}$ is defined as pointing in the opposite direction, then ϕ reverses its sign.

Apparently, $\mathcal{L}$ lies in the bisector plane of segment $\overline{PP_F}$. Hence, to find $\mathcal{L}$, all we need is two points in their original and final positions. The intersection of the two bisector planes of the segments defined by these points then gives $\mathcal{L}$. We thus start by finding the bisector planes Π_A and Π_B of segments $\overline{AA_F}$ and $\overline{BB_F}$, respectively. To this end, we recall eq.(3.2), thus obtaining

$$\Pi_A : \quad (\boldsymbol{a}_F - \boldsymbol{a})^T \boldsymbol{p} + \frac{1}{2}(||\boldsymbol{a}||^2 - ||\boldsymbol{a}_F||^2) = 0$$

$$\Pi_B : \quad (\boldsymbol{b}_F - \boldsymbol{b})^T \boldsymbol{p} + \frac{1}{2}(||\boldsymbol{b}||^2 - ||\boldsymbol{b}_F||^2) = 0$$

In our case,

$$||\boldsymbol{a}||^2 = ||\boldsymbol{a}_F||^2 = d_A^2, \quad ||\boldsymbol{b}||^2 = ||\boldsymbol{b}_F||^2 = d_B^2$$

$$\boldsymbol{a}_F - \boldsymbol{a} = \begin{bmatrix} -d_A \\ 0 \\ d_A \end{bmatrix}, \quad \boldsymbol{b}_F - \boldsymbol{b} = \begin{bmatrix} d_B \\ -d_B \\ 0 \end{bmatrix}$$

Hence, the equations of the two planes are

$$\Pi_A \quad : \quad -d_A x + d_A z = 0$$

$$\Pi_B \quad : \quad d_B x - d_B y = 0$$

which gives a homogeneous system of two linear equations in three unknowns. These two equations thus define a line $\mathcal{L}$ passing through the origin—the coordinates of the origin, (0,0,0), verify the two equations—which is nothing but the axis of rotation sought.

To compute the direction cosines of $\mathcal{L}$, (λ, μ, ν), all we need is substitute

$$\lambda \leftarrow x, \quad \mu \leftarrow y, \quad \nu \leftarrow z$$

in the two above equations, and impose the condition that the sum of the squares of the three direction cosines be unity, i.e.,

$$\begin{aligned} -\lambda + \nu &= 0 \\ \lambda - \mu &= 0 \\ \lambda^2 + \mu^2 + \nu^2 &= 1 \end{aligned}$$

From the first and the second equations, we obtain

$$\nu = \lambda, \quad \mu = \lambda$$

whence the third equation yields

$$3\lambda^2 = 1 \quad \Rightarrow \lambda = \pm\frac{\sqrt{3}}{3}$$

Picking up the positive sign, for example, we obtain

$$\lambda = \mu = \nu = \frac{\sqrt{3}}{3} \quad \Rightarrow \boldsymbol{e} = \frac{\sqrt{3}}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

the last equation following from the realization that the direction cosines of $\mathcal{L}$ are nothing but the components of the unit vector $\boldsymbol{e}$, giving the direction of $\mathcal{L}$, thereby finding the axis of rotation. Further, to find the angle of rotation, we need a point Q on $\mathcal{L}$ that lies on the normal to $\mathcal{L}$ from, say A—B might as well be taken, without affecting the final result—so that $\overline{QA}$ is normal to $\mathcal{L}$. If we let (ξ, η, ζ) be the coordinates of Q, then the foregoing normality condition can be expressed as

$$\boldsymbol{e}^T(\boldsymbol{q} - \boldsymbol{a}) = 0 \quad \Rightarrow \frac{\sqrt{3}}{3}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} \xi - d_A \\ \eta \\ \zeta \end{bmatrix} = 0$$

where $\boldsymbol{q}$ is the position vector of Q. Hence,

$$\xi - d_A + \eta + \zeta = 0$$

thereby obtaining one equation for the three coordinates of Q. Two more equations are available if we realize that Q is a point of $\mathcal{L}$, and hence, its coordinates verify the equations of Π_A and Π_B:

$$-\xi + \zeta = 0, \quad \xi - \eta = 0$$

The two above equations yield $\eta = \xi$, $\zeta = \xi$. When these expressions are substituted in the foregoing normality condition, we obtain

$$3\xi = d_A \quad \Rightarrow \xi = \frac{1}{3}d_A$$

Hence,

$$\boldsymbol{q} = \left[\begin{array}{ccc} d_A/3 & d_A/3 & d_A/3 \end{array}\right]^T$$

Now, ϕ can be obtained from the relation

$$(\boldsymbol{a} - \boldsymbol{q})^T(\boldsymbol{a}_F - \boldsymbol{q}) = ||\boldsymbol{a} - \boldsymbol{q}||^2 \cos\phi$$

where

$$\boldsymbol{a} - \boldsymbol{q} = \begin{bmatrix} 2d_A/3 \\ -d_A/3 \\ -d_A/3 \end{bmatrix}, \quad \boldsymbol{a}_F - \boldsymbol{q} = \begin{bmatrix} -d_A/3 \\ -d_A/3 \\ 2d_A/3 \end{bmatrix}, \quad ||\boldsymbol{a} - \boldsymbol{q}||^2 = \frac{2}{3}d_A^2$$

Therefore,

$$\frac{d_A^2}{q}(-2 + 1 - 2) = \frac{2}{3}d_A^2 \cos\phi, \quad \Rightarrow \cos\phi = -\frac{1}{2}$$

from which $\phi = 120°$ or $240°$. To destroy the ambiguity, we take into account the direction given by $\boldsymbol{e}$. Hence,

$$(\boldsymbol{a} - \boldsymbol{q}) \times (\boldsymbol{a}_F - \boldsymbol{q}) = (\sin\phi)||\boldsymbol{a} - \boldsymbol{q}||^2\boldsymbol{e}$$

If we dot-multiply both sides of the above equation by $\boldsymbol{e}$, we obtain an equation for $\sin\phi$:

$$||\boldsymbol{a} - \boldsymbol{q}||^2 \sin\phi = (\boldsymbol{a} - \boldsymbol{q}) \times (\boldsymbol{a}_F - \boldsymbol{q}) \cdot \boldsymbol{e} = \det([\boldsymbol{a} - \boldsymbol{q}, \boldsymbol{a}_F - \boldsymbol{q}, \boldsymbol{e}])$$

Therefore,

$$\frac{2}{3}d_A^2 \sin\phi = \det\begin{bmatrix} \frac{2d_A}{3} & \frac{-d_A}{3} & \frac{\sqrt{3}}{3} \\ \frac{-d_A}{3} & \frac{-d_A}{3} & \frac{\sqrt{3}}{3} \\ \frac{-d_A}{3} & \frac{2d_A}{3} & \frac{\sqrt{3}}{3} \end{bmatrix}$$

If we now recall relation (1.76), the above determinant simplifies to

$$\frac{2}{3}d_A^2 \sin\phi = \frac{1}{3^3}\det\begin{bmatrix} 2d_A & -d_A & \sqrt{3} \\ -d_A & -d_A & \sqrt{3} \\ -d_A & 2d_A & \sqrt{3} \end{bmatrix}$$

Upon expansion of the determinant by cofactors of its third column, we obtain

$$\begin{aligned}
\frac{2}{3}d_A^2 \sin\phi &= \frac{\sqrt{3}}{3^3}[\det\begin{bmatrix} -d_A & -d_A \\ -d_A & 2d_A \end{bmatrix} - \det\begin{bmatrix} 2d_A & -d_A \\ -d_A & 2d_A \end{bmatrix} + \det\begin{bmatrix} 2d_A & -d_A \\ -d_A & -d_A \end{bmatrix}] \\
&= \frac{\sqrt{3}}{3^3}[(-2d_A^2 - d_A^2) - (4d_A^2 - d_A^2) + (-2d_A^2 - d_A^2)] \\
&= \frac{\sqrt{3}}{3^3}(-9d_A^2) = -\frac{\sqrt{3}}{3}d_A^2
\end{aligned}$$

whence

$$\sin\phi = -\frac{\sqrt{3}}{2}$$

which verifies $\sin^2\phi + \cos^2\phi = 1$ with the result obtained above for $\cos\phi$. We can thus conclusively state that, for the above value of $\boldsymbol{e}$, $\phi = 240°$.

4.3.4 Reflection

A reflection, similar to a rotation, preserves the distance between every two points of an object, but changes its *hand*. For example, making abstraction of the internal organs, the human body can be regarded as a symmetric object, its plane of symmetry being the *sagital plane*. This plane divides the body into two symmetric parts, left and right. The left part is a reflection of the right part, the sagital plane thus being the *plane of reflection*.

A three-dimensional reflection (mirroring) is usually obtained by coordinate transformations about a specified reflection plane.

- The matrix representing a reflection about the plane $x = 0$ is given by

$$\boldsymbol{M}_X = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.33}$$

- The matrix representing a reflection about the plane $y = 0$ is given by

$$\boldsymbol{M}_Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.34}$$

- The matrix representing a reflection about the plane $z = 0$ is given by

$$\boldsymbol{M}_Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{4.35}$$

To reflect an object about any arbitrary plane, combined transformations involving rotation and reflections will have to be performed.

4.4 Computer Implementation of 3D Affine Transformations

Similar to the case of affine transformations in 2D, the computer implementation of affine transformations in 3D calls for multiplying the position vector $\boldsymbol{p}$ in homogeneous coordinates by the *inverse* of the desired transformation, thereby obtaining a new vector $\boldsymbol{q}$ in the form

$$\boldsymbol{q} = \boldsymbol{T}^{-1}\boldsymbol{p} \tag{4.36}$$

For example, let P be an arbitrary point of the solar panel of Fig. 4.16(a), of homogeneous coordinates arrayed in vector $\boldsymbol{p}$, as given by eq.(4.22b). This point is carried into a point Q of the same solar panel in the orientation displayed in Fig. 4.16(d), of homogeneous coordinates arrayed in vector $\boldsymbol{q}$. The homogeneous-transformation matrix carrying $\boldsymbol{p}$ into $\boldsymbol{q}$ can be readily derived if we: (a) recall expression (1.91) for the inverse of a 4×4 homogeneous-transformation matrix; realize that the transformation is a pure rotation, and hence, $\boldsymbol{t} = \mathbf{0}$; and (b) notice that matrix $\boldsymbol{M}$, derived in Subsection 4.3.3, is orthogonal and hence, its inverse is simply its transpose. We thus obtain:

$$\boldsymbol{T}^{-1} = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{q} = \begin{bmatrix} -z \\ y \\ x \\ 1 \end{bmatrix} \tag{4.37}$$

4.5 Techniques for 3D Object Modelling

Three-dimensional objects can be regarded as regions of space bounded by closed surfaces. In this section we study the various techniques available for the production of such surfaces.

4.5.1 Surfaces of revolution

A simple family of surfaces is obtained by rotating a plane curve around an axis, thereby obtaining a surface of revolution.

For example, a circular cylinder is formed by rotating a line segment parallel to the Z-axis through an angle of 2π around the same Z-axis.

We will describe here the generation of surfaces of revolution by means of the rotation of a plane curve Γ in the XZ plane, the *generatrix*, about the Z-axis. Shown in Fig. 4.19 is the generatrix Γ and the displacement of one arbitrary point of Γ upon a rotation of Γ about Z through an angle θ.

The homogeneous coordinates of P and P' are stored in the 4-dimensional arrays $\boldsymbol{p}$ and $\boldsymbol{p'}$ which are related by an affine transformation of the form of eq.(4.22a), with $\boldsymbol{M}$ representing a rotation about the Z-axis through an angle θ, namely

$$\boldsymbol{M} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For the case depicted in Fig. 4.19, we have, in homogeneous-coordinate form,

$$\boldsymbol{p} = \begin{bmatrix} x \\ 0 \\ z \\ 1 \end{bmatrix}, \quad \boldsymbol{p'} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ 0 \\ z \\ 1 \end{bmatrix}$$

i.e.,

$$\boldsymbol{p'} = \begin{bmatrix} x\cos\theta \\ x\sin\theta \\ z \\ 1 \end{bmatrix}$$

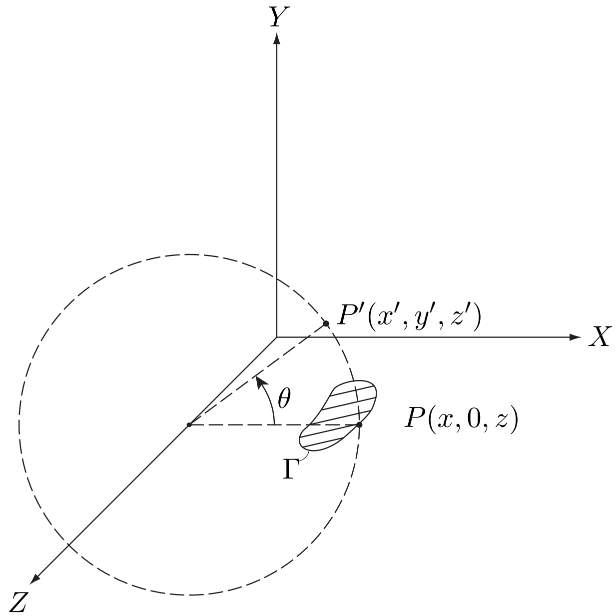


Figure 4.19: Generation of a surface of revolution by means of the rotation of a generatrix Γ in the XZ plane about the Z-axis

Example 4.5.1 *Construct an O-ring of cross-section of radius r and main radius $a > r$.*

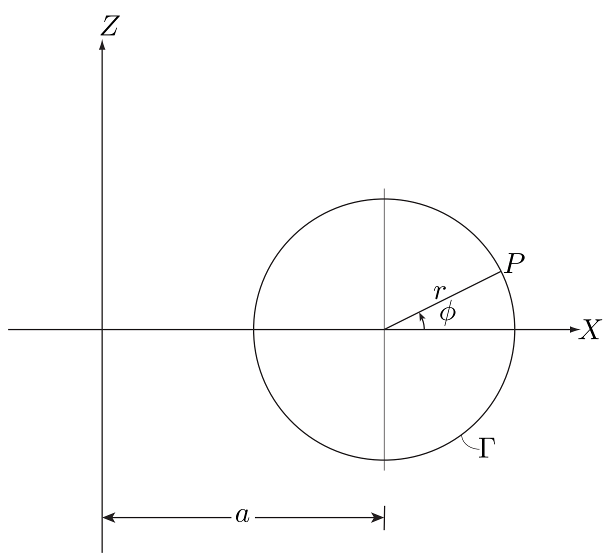


Figure 4.20: Construction of an O-ring by revolving the circle Γ about the Z-axis

Solution: The O-ring is generated by revolving the circle Γ lying in the XZ plane, as depicted in Fig. 4.20, about the Z-axis. Representing the circle in polar coordinates, we have

$$\boldsymbol{p} = \begin{bmatrix} a + r\cos\phi \\ 0 \\ r\sin\phi \\ 1 \end{bmatrix} \Rightarrow \boldsymbol{p'} = \begin{bmatrix} (a + r\cos\phi)\cos\theta \\ (a + r\cos\phi)\sin\theta \\ r\sin\phi \\ 1 \end{bmatrix}$$

A small piece of code was written using computer-algebra software [5] to produce a rendering of the O-ring, namely

```
> restart; with(plots):
> with(LinearAlgebra):
> R:=<<cos(theta),sin(theta),0,0>|<sin(theta),cos(theta),0,0>|
> <0,0,1,0>|<0,0,0,1>>;
```

$$\begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

```
> p:=<a+r*cos(phi),0,r*sin(phi),1>;
```

$$\begin{bmatrix} 3 + \cos(\phi) \\ 0 \\ \sin(\phi) \\ 1 \end{bmatrix}$$

```
> Rp:=R.p;
```

$$\begin{bmatrix} \cos(\theta)(3 + \cos(\phi)) \\ \sin(\theta)(3 + \cos(\phi)) \\ \sin(\phi) \\ 1 \end{bmatrix}$$

```
> a:=3; r:=1;
```

$$3$$
$$1$$

```
> plot3d(Rp[1..3], theta=0..2*Pi,
> phi=0..2*Pi, scaling=constrained,grid=[60,60]);
```

The code produced the rendering of Fig. 4.5.1, with the numerical values $r = 10$mm, $a = 50$mm.

4.5.2 Extrusion

Extrusion is a procedure by which a surface is generated through the movement of a line, a curve segment, a polygon, and so forth, i.e., a generatrix, along a defined path. The paths followed in extrusion operations can be straight lines or curves. We will focus on extrusion along lines. The corresponding extruded surface is represented in parametric form as:

[5] Maplesoft's Maple 10.

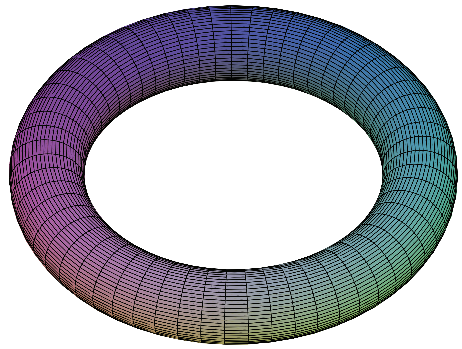


Figure 4.21: Computer rendering of an O-ring with $r = 10$ mm, $a = 50$ mm

$$\boldsymbol{q}(t, s) = \boldsymbol{T}(s)\boldsymbol{p}(t) \tag{4.38}$$

where $\boldsymbol{p}(t)$ is the 4-dimensional array of homogeneous coordinates of a point P of the generatrix, in parametric form, and $\boldsymbol{T}(s)$ is the extrusion transformation based on the shape of the path, given in terms of a second parameter, s, and $\boldsymbol{q}(t, s)$ is the 4-dimensional array of homogeneous coordinates of the transformed point Q.

The extrusion transformation can contain translations, scalings, rotations, or combinations of these transformations. For the case in which the path is a line, all the points of the generatrix Γ, which we will assume to be a planar curve in the XZ plane, translate in the direction of extrusion, given by a unit vector $\boldsymbol{e}$.

The displacement of every point P of Γ is thus $s\boldsymbol{e}$, where $s \geq 0$ is the translation parameter, the extrusion matrix then taking the form

$$\boldsymbol{T}(s) = \begin{bmatrix} \boldsymbol{M} & s\boldsymbol{e} \\ \boldsymbol{0}^T & 1 \end{bmatrix}$$

Matrix $\boldsymbol{M}$ can be constant or a function of s, depending on the type of extrusion at hand. Moreover, while we have assumed that Γ lies in the XZ plane, we need not impose any constraint on $\boldsymbol{e}$, except that it is a unit vector. A few examples will illustrate the power of the extrusion transformation to construct a variety of solids.

Example 4.5.2 *A shaft of radius r and length l can be constructed by the* simple extrusion *of a circle centred at the origin of the XZ plane, with direction of extrusion given by the Y-axis. In this case,*

$$\boldsymbol{p}(t) = \begin{bmatrix} r\cos t \\ 0 \\ r\sin t \\ 1 \end{bmatrix}, \quad \boldsymbol{e} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{M} = \boldsymbol{1}, \quad 0 \leq s \leq l$$

Whence,

$$\boldsymbol{q}(t,s) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & s \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r\cos t \\ 0 \\ r\sin t \\ 1 \end{bmatrix} = \begin{bmatrix} r\cos t \\ s \\ r\sin t \\ 1 \end{bmatrix}$$

Example 4.5.3 *A tapered shaft of largest radius r, smallest radius $r/2$, length l and tapering angle α can be constructed using the same generatrix Γ and the same direction of extrusion as in Example 4.5.2. The difference now is that $\boldsymbol{M}$ involves a scaling by the angle of tapering, so that*

$$\boldsymbol{M} = k(s)\mathbf{1}$$

where $k(s)$ is a scaling factor, which is determined with the aid of Fig. 4.22. Notice that α can be computed from the dimensions r and l.

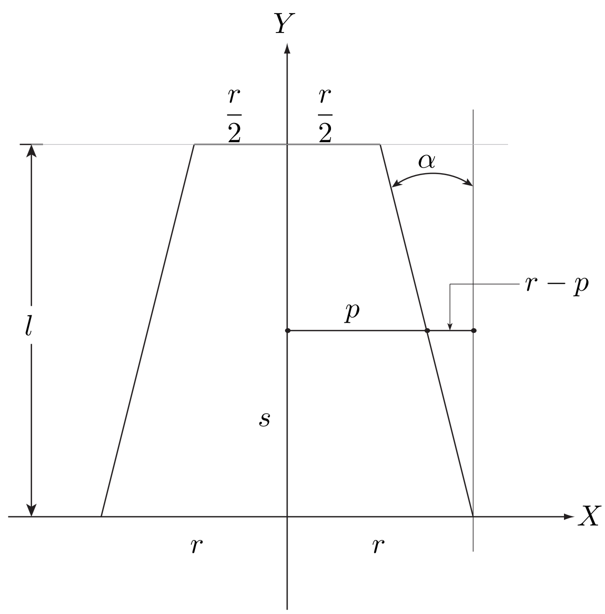


Figure 4.22: Construction of a tapered shaft

The reader should be able to verify that the scaling factor $k(s)$ is given by

$$k(s) \equiv \frac{p}{r} = 1 - \frac{s}{2l}$$

A three-dimensional rendering of the shaft, for $l = 300$ mm and $r = 60$ mm is included in Fig. 4.23.

Example 4.5.4 *Construction of a screw. We show in Fig. 4.24(a) a sketch of a common type of screw, illustrating its terminology. In this sketch, the vee-shaped crests and roots serve purposes of sketch simplicity; in practice, these are either flattened or rounded, as illustrated in Fig. 4.24(b). Devise a means of constructing the threaded portion of the screw. For simplicity, keep the vee-shapes.*

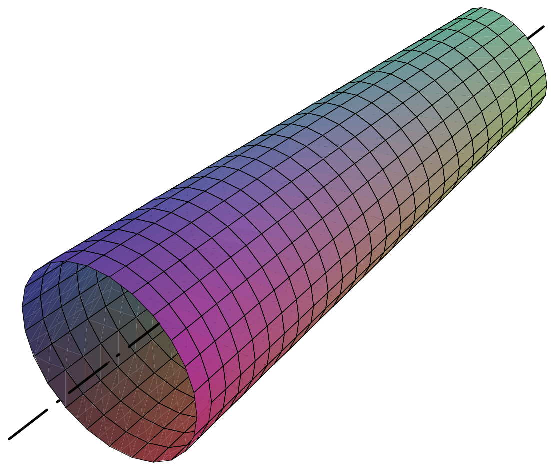


Figure 4.23: Three-dimensional rendering of the shaft, with $l = 300$ mm and $r = 60$ mm

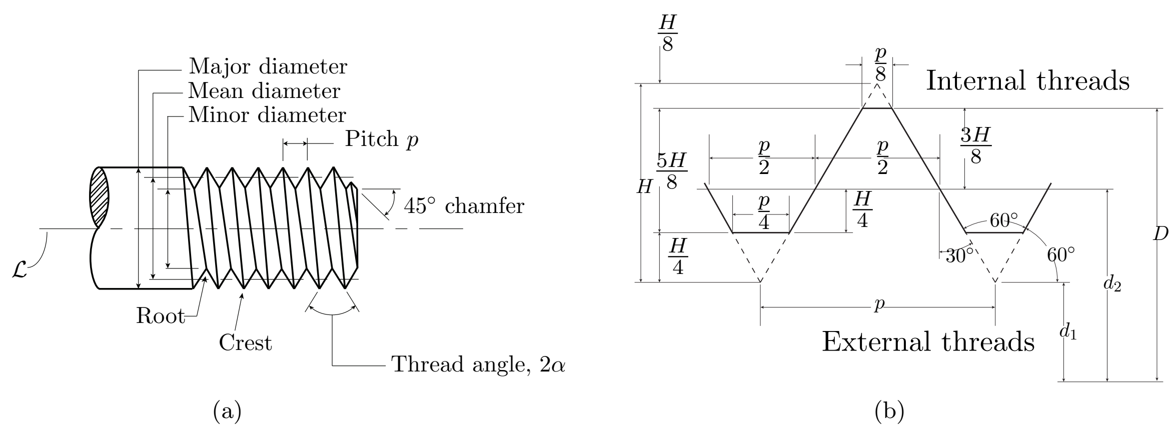
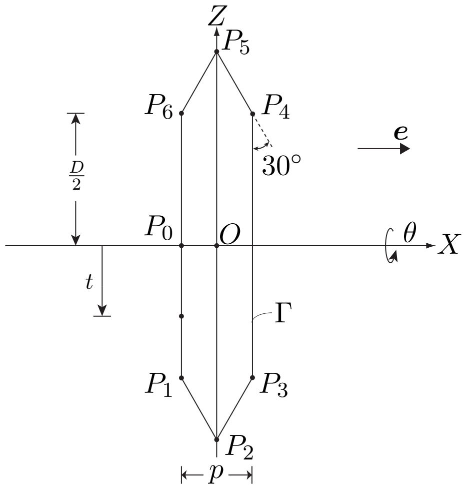


Figure 4.24: The geometry of a common type of screw: (a) terminology; (b) flattening of the crests and roots for metric M and MJ threads, with $p =$ pitch



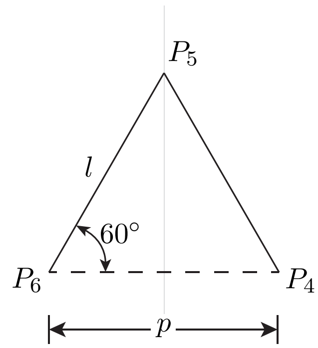


Figure 4.25: The generatrix Γ for the construction of the threaded surface of a screw: (a) general layout; (b) detail of the vee-shaped parts

Solution: We shall use extrusion along the direction of the axis $\mathcal{L}$ of the screw combined with rotation about the same axis, the generatrix Γ being illustrated in Fig. 4.25(a).

In order to represent Γ parametrically, we use the length t along the profile. The lengths of the vertical segments of Γ are straightforward, those of the inclined segments, of length l, being derived from the detail in Fig. 4.25(b): Because l and p are sides of the same equilateral triangle,

$$l = p$$

Hence, Γ is described, parametrically, as

$$\begin{array}{lll}
0 \le t \le \frac{D}{2}: & x = -\frac{p}{2}, & z = -t \\
\frac{D}{2} \le t \le \frac{D}{2} + p: & x = -\frac{p}{2} + (t - \frac{D}{2}) \cos 60^\circ, & z = -\frac{D}{2} + (-t + \frac{D}{2}) \sin 60^\circ \\
\frac{D}{2} + p \le t \le \frac{D}{2} + 2p: & x = (t - \frac{D}{2} - p) \cos 60^\circ, & z = -\frac{D}{2} + (t - \frac{D}{2} - 2p) \sin 60^\circ \\
\frac{D}{2} + 2p \le t \le \frac{3D}{2} + 2p: & x = \frac{p}{2}, & z = t - D - 2p \\
\frac{3D}{2} + 2p \le t \le \frac{3D}{2} + 3p: & x = \frac{p}{2} + (-t + \frac{3D}{2} + 2p) \cos 60^\circ, & z = \frac{D}{2} + (t - \frac{3D}{2} + 2p) \sin 60^\circ \\
\frac{3D}{2} + 3p \le t \le \frac{3D}{2} + 4p: & x = (-t + \frac{3D}{2} + 3p) \cos 60^\circ, & z = \frac{D}{2} + (-t + \frac{3D}{2} + 4p) \sin 60^\circ \\
\frac{3D}{2} + 3p \le t \le 2D + 4p: & x = -\frac{p}{2}, & z = -t + 2D + 4p)
\end{array}$$

With Γ available in parametric form, all we need is $\boldsymbol{t}$ and $\boldsymbol{M}$; $\boldsymbol{t}$ is a translation in the direction of $\boldsymbol{e}$, whence,

$$\boldsymbol{t} = s\boldsymbol{e}, \quad \boldsymbol{e} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Moreover, $\boldsymbol{M}$ is a rotation about X of angle $\theta = s/p$, whence, recalling eq.(4.32),

$$\boldsymbol{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(s/p) & -\sin(s/p) \\ 0 & \sin(s/p) & \cos(s/p) \end{bmatrix}$$

thereby completing the affine transformation of interest. The transformation was implemented using computer algebra in the piece of code displayed below:

```
>  restart:  with(plots):
>  # definition of the generatrix
>  Gamma
>  pitch:=1.5:  d:=10:  # Note:
>  Maple reserves "D" for derivative; we use "d" to represent "D" in
>  the definition of the generatrix
>  x1:= piecewise(t<=d/2, -pitch/2,
>                 t<=d/2+pitch,-pitch/2+(t-d/2)*cos(Pi/3),
>                 t<=d/2+2*pitch,(t-d/2-pitch)*cos(Pi/3),
>                 t<=3*d/2+2*pitch, pitch/2,
>                 t<=3*d/2+3*pitch,
>  pitch/2+(-t+3*d/2+2*pitch)*cos(Pi/3),
>                 t<=3*d/2+4*pitch, (-t-3*d/2+3*pitch)*cos(Pi/3),
>                 t<=2*d+4*pitch,-pitch/2):
>  z1:= piecewise(t<=d/2, -t,
>                 t<=d/2+pitch, -d/2+(-t+d/2)*sin(Pi/3),
>                 t<=d/2+2*pitch, -d/2+(t-d/2-2*pitch)*sin(Pi/3),
>                 t<=3*d/2+2*pitch, -d+t-2*pitch,
>                 t<=3*d/2+3*pitch, d/2+(t-3*d/2-2*pitch)*sin(Pi/3),
>                 t<=3*d/2+4*pitch,d/2+(-t+3*d/2+4*pitch)*sin(Pi/3),
>                 t<=2*d+4*pitch,-t+2*d+4*pitch)):
```

The piece of code below yielded the rendering shown in Fig. 4.26, with parameters $D = 10$ mm and $p = 1.5$ mm.

```
>  #visualization of the generatrix
>  plot([x1,z1,t=0..2*d+4*pitch],scaling=constrained);
>  # definition of the transformation
>  matrix T
>  T:=Matrix([[1,0,0,s/(2*pitch)],[0,cos(s/pitch),-sin(s/pitch),0],
>  [0,sin(s/pitch),cos(s/pitch),0]]);
>  p:=Vector([x1,0,z1,1]):
>  q:=evalm(T.p):
>  # screw rendering
>  plot3d([q[1],q[2],q[3]],t=0..2*d+4*pitch,s=0..30,scaling=constrained);
```

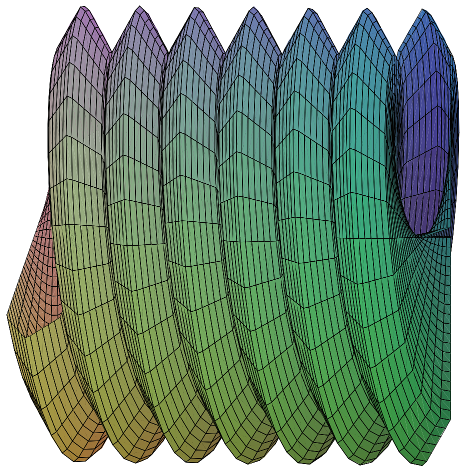


Figure 4.26: Computer rendering of a coarse-pitch screw, with $D = 10$ mm and $p = 1.5$ mm

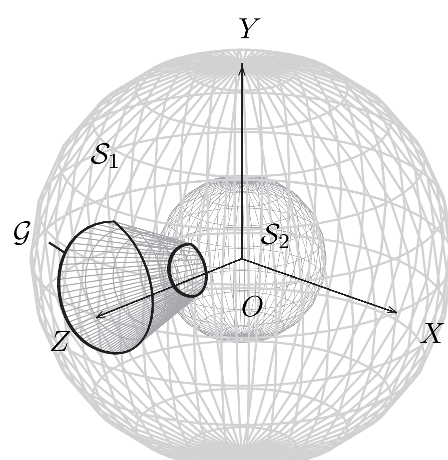


Figure 4.27: Conic extrusion

Conic extrusion

Consider a closed curve[6] $\mathcal{G}$ lying on a sphere $\mathcal{S}_1$ of centre O. A *conic extrusion* is a transformation yielding a *conic surface* with generatrix $\mathcal{G}$ and vertex O. The surface is generated by $\mathcal{G}$ as the sphere $\mathcal{S}_1$ is scaled to a concentric sphere $\mathcal{S}_2$, as depicted in Fig. 4.27. Notice that $\mathcal{S}_2$ can be either smaller than $\mathcal{S}_1$, as in Fig. 4.27, in which case we have an *inward extrusion*, or larger than $\mathcal{S}_1$, in which case we have an *outward extrusion*. In fact, the conic extrusion is a particular case of a 3D scaling, in which all three factors are identical, i.e., an isotropic transformation, its homogeneous transformation matrix $\boldsymbol{T}_{ce}$ being given by

$$\boldsymbol{T}_{ce} = \begin{bmatrix} s\mathbf{1}_{3\times 3} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} s & 0 & 0 & 0 \\ 0 & s & 0 & 0 \\ 0 & 0 & s & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where s is the unique scaling factor.

Conic extrusions find applications in the design of spherical domes, bevel gears and spherical cam mechanisms.

4.5.3 Free-form Surfaces

As in the case of curves, some surfaces cannot be totally described by simple formulas. Among these are surfaces used in the design of automobile bodies, ship hulls, aircraft wings, and so forth. These surfaces are usually described by a series of "patches", in the same way that a patchwork quilt is put together. The free-form curve tools, Bézier curves, B-splines, etc., can be used in free-form surface design.

4.6 CAD Tools for Creating 3D Objects

Surfaces can be created using a number of different techniques supported by CAD software. The technique used is determined both by the shape to be created and by the tools available in the CAD surface modeller at one's disposal. Among the most popular methods for creating surfaces, we can cite *extrusion* and *revolution*. In Section 4.5, we studied the homogeneous transformations involved in the construction of extruded objects and objects of revolution. Here we expand on the capabilities of CAD software for these tasks.

As illustrated in Fig. 4.28, in extrusion operations, the directrix is typically a planar curve, while the generatrix can be a line, a planar curve, or a 3D curve.

Many features in a model may be created through the use of extrusion operations. Most CAD systems use methods of automating object generation. In an extrusion operation, a closed polygon, called a profile, is drawn on a plane and is moved or swept along a defined path for a defined length.

Figure 4.29 gives examples of extrusion along a line.

As shown in Fig. 4.30, the distance through which a profile is swept can be determined in a number of ways, including: *blind*, *through-all*, or *to-next*.

The extruded feature will either add or subtract material from the existing model, depending on how the feature has been defined, as illustrated in Figure 4.31.

[6]Apparently, $\mathcal{G}$ **cannot** be a planar contour, unless, of course, the contour is a circle.

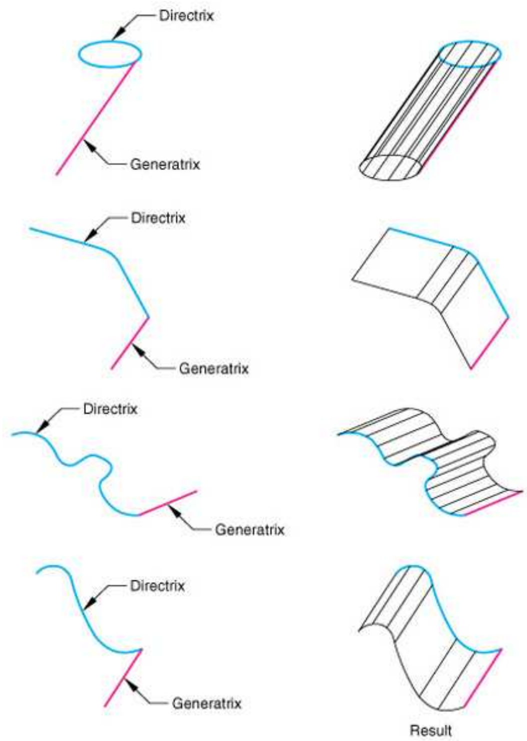


Figure 4.28: How to generate extruded surfaces.

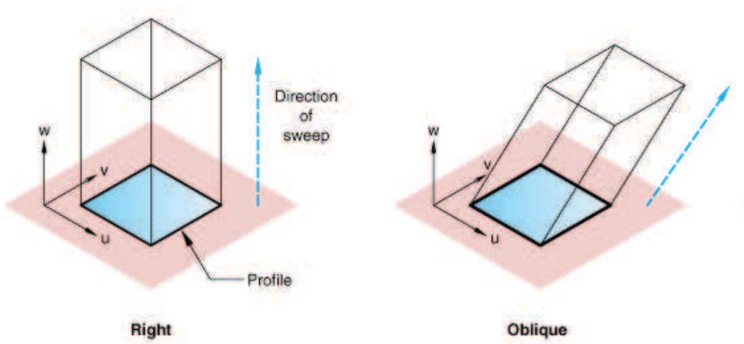


Figure 4.29: Types of extrusion along a line

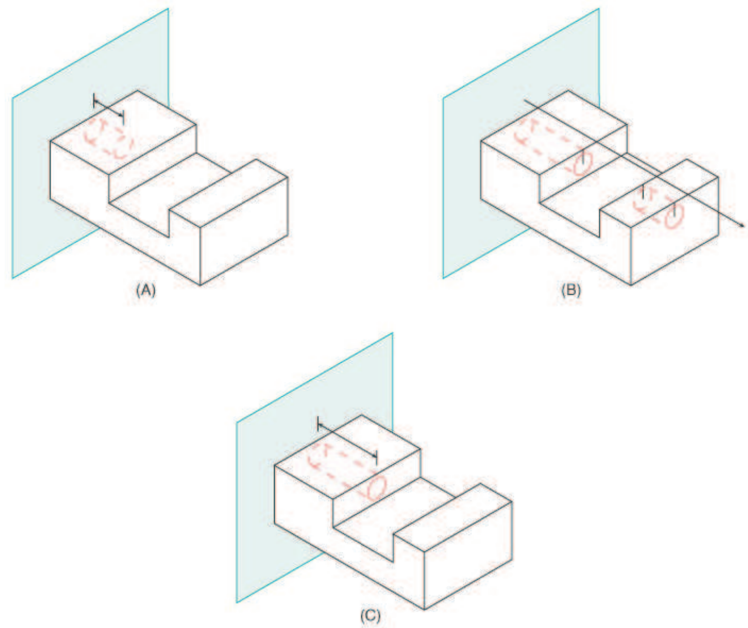


Figure 4.30: Defining an extrusion distance: (A) blind; (B) through-all; and (C) to-next

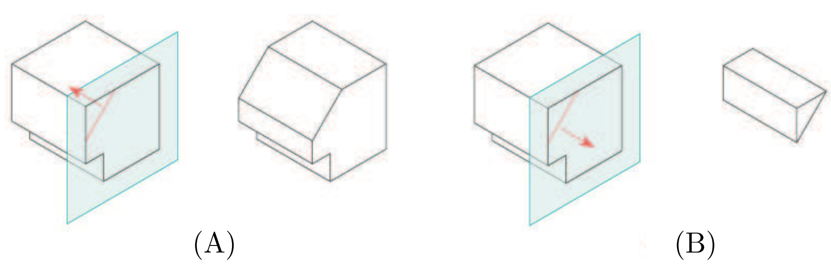


Figure 4.31: Determining the removal side of an extrusion

It is possible to create more complex solid models using a combination of extrusion and Boolean operations, as shown in Fig. 4.32.

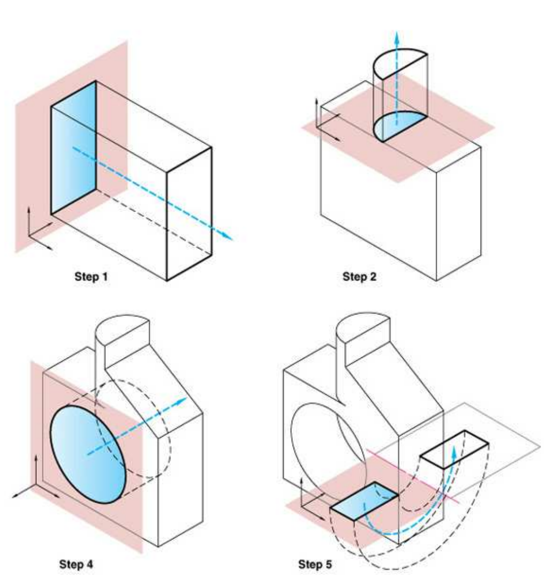


Figure 4.32: Creating a solid model using extrusion and Boolean operations

Chapter 5

Multi-Visualization

5.1 View of Part Model

The techniques used for viewing 3D models are based on the principles of projection theory. The computer screen, like a sheet of paper, is two-dimensional. Therefore, 3D forms must be projected onto 2D. Recall that the primary elements in creating a projection are the model (object), the viewer, and an image (view) plane, as illustrated in Fig. 5.1. A coordinate system is attached to each of these elements and is used to define the spatial relationship between the elements.

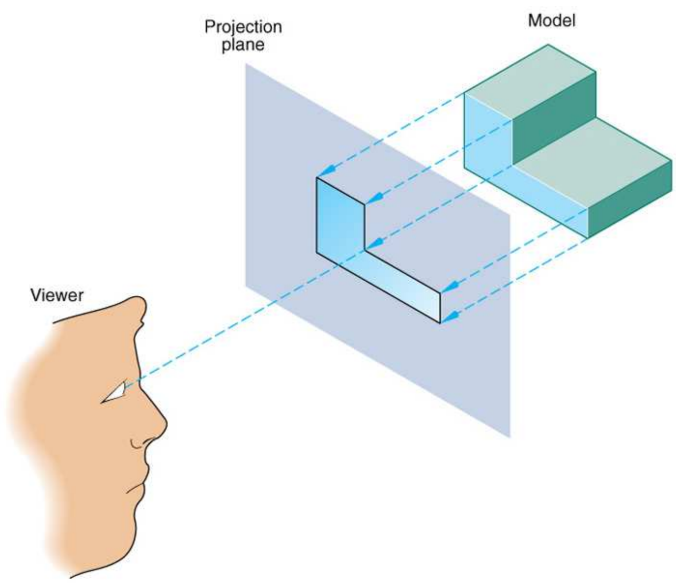


Figure 5.1: Elements of a projection system

The view camera, as shown in Fig. 5.2, is a metaphor used to describe the viewing process with respect to 3D models in various CAD systems. For each view, there is a camera, and there is an image plane onto which the model is projected. The camera records the image on the plane and broadcasts that image to the computer screen. The broadcast image is contained within a viewport on the screen; viewports may be resizable and relocatable, or fixed, depending on the system.

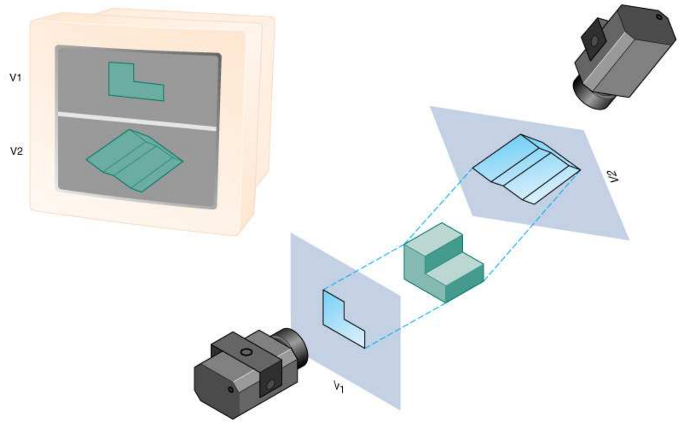


Figure 5.2: The view camera

5.2 Projections

The problem of projecting a three-dimensional object onto a two-dimensional surface has been studied by engineers, architects, and artists for centuries. Computer graphics systems also address problems related to projections.

Planar geometric projections, of most interest to engineers, can be classified as shown in Fig. 5.3. In planar projections, a viewing direction is established from the observer to the object by means of projector lines that cut through a plane where the projection appears.

Projection methods are developed along two lines: *perspective* and *parallel*. Projection theory comprises the principles used to represent graphically 3D objects and structures on 2D media. Drawing more than one face of an object by moving your line of sight relative to the object helps in understanding the 3D form. A line of sight is an imaginary ray of light between an observer's eye and the object.

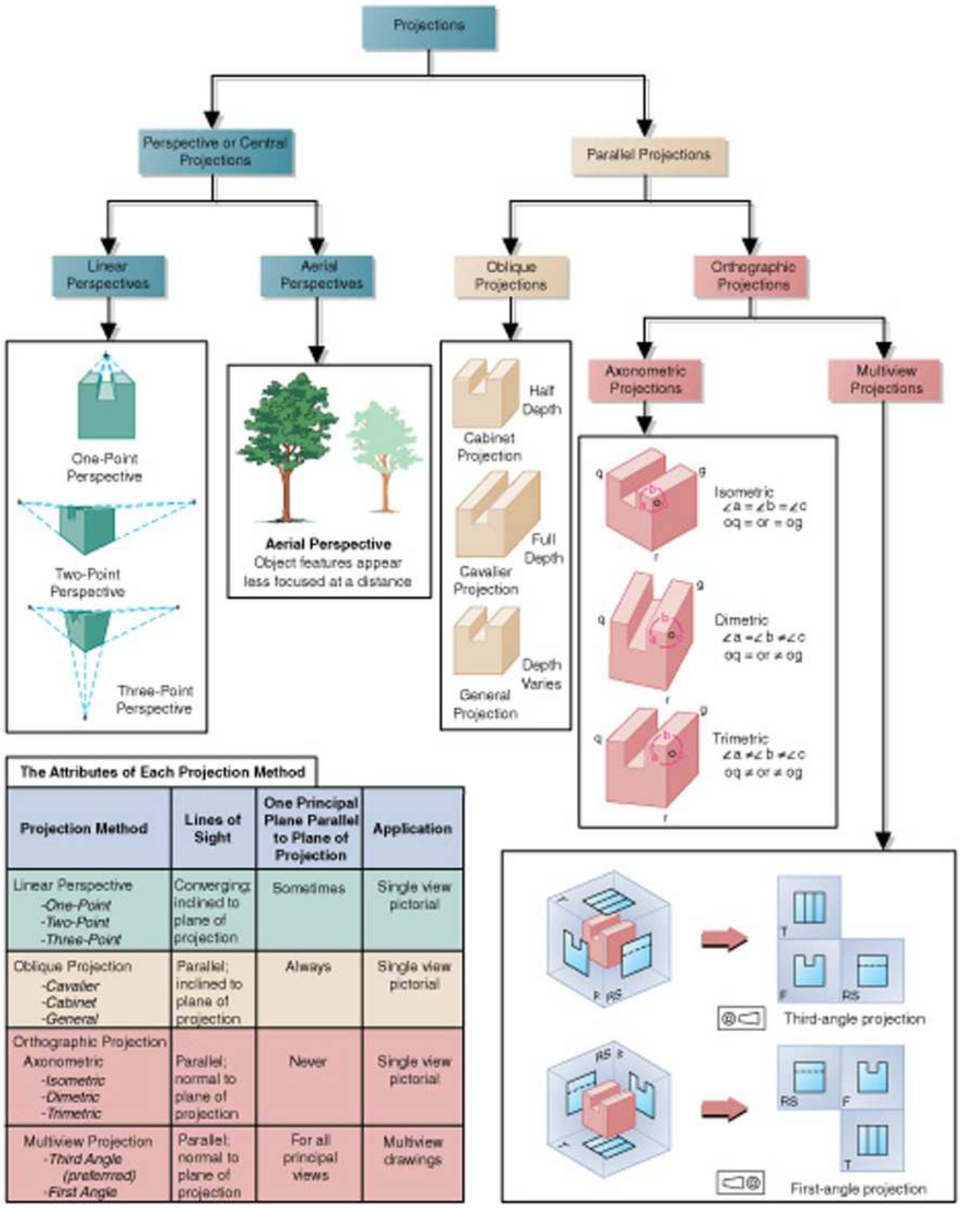
In **perspective projection**, all lines of sight start at a single point, and the object is positioned at a finite distance and viewed from a single point, as illustrated in Fig. 5.4.

In **parallel projection**, as shown in Fig. 5.5, all lines of sight are parallel, the object is positioned at infinity and viewed from multiple points on an imaginary line parallel to the object. The 3D object is transformed into a 2D representation or a plane of projection that is an imaginary flat plane upon which the image created by the lines of sight is projected. The paper or computer screen on which the graphic is created is a plane of projection.

5.2.1 Multiview orthographic projections

Orthographic projection is a parallel projection technique in which the plane of projection is positioned between the observer and the object, and is perpendicular to the parallel lines of sight, as illustrated in Fig. 5.6. Orthographic projection techniques can be used to produce both pictorial and multiview drawings; they are commonly used in engineering.

Consider a point $P(x, y, z)$, projected onto the XY-plane. The projection of point P is $P'(x', y', 0)$; the homogeneous transformation that produces this projection can be expressed



Projection Method	Lines of Sight	One Principal Plane Parallel to Plane of Projection	Application
Linear Perspective -One-Point -Two-Point -Three-Point	Converging; inclined to plane of projection	Sometimes	Single view pictorial
Oblique Projection -Cavalier -Cabinet -General	Parallel; inclined to plane of projection	Always	Single view pictorial
Orthographic Projection Axonometric -Isometric -Dimetric -Trimetric	Parallel; normal to plane of projection	Never	Single view pictorial
Multiview Projection -Third Angle (preferred) -First Angle	Parallel; normal to plane of projection	For all principal views	Multiview drawings

Figure 5.3: The different projection techniques

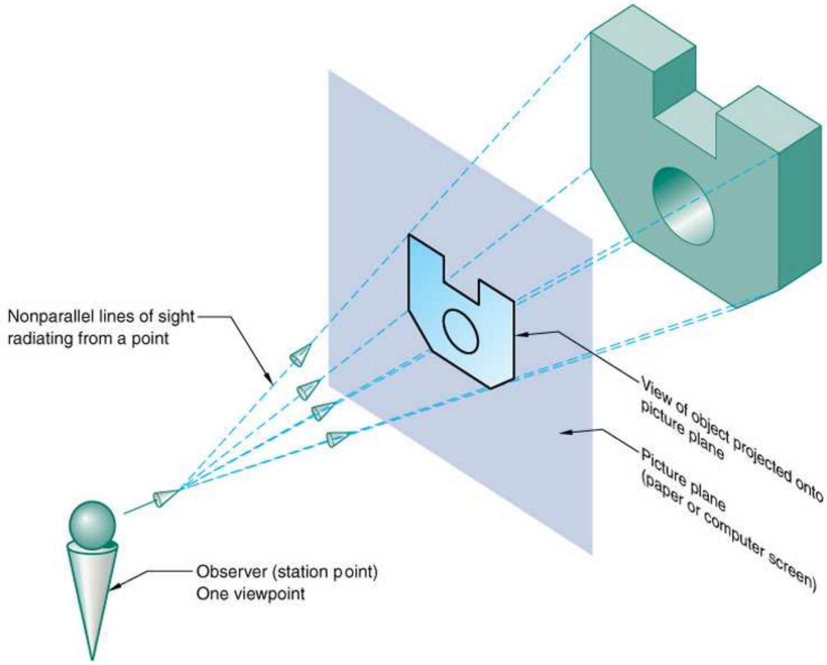


Figure 5.4: Perspective projection

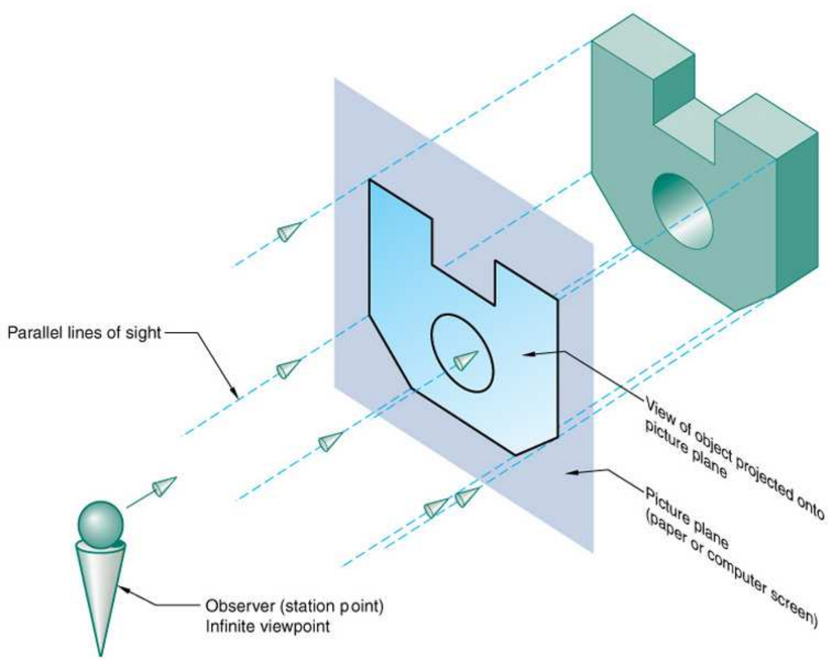


Figure 5.5: Parallel projection

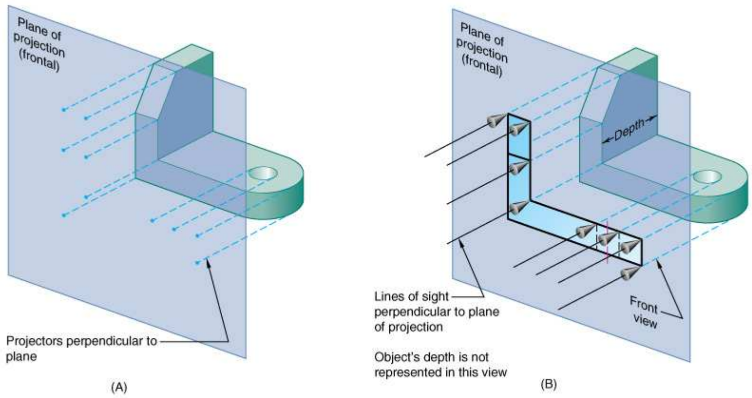


Figure 5.6: Orthographic projection

as:

$$\begin{bmatrix} x' \\ y' \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{5.1}$$

Therefore, the projection matrix is:

$$\boldsymbol{P}_{XY} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.2}$$

The same approach would be used for the orthographic projection onto the XZ or YZ planes. The respective projection matrices would be:

$$\boldsymbol{P}_{XZ} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{P}_{YZ} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.3}$$

This type of projection can also be obtained by simply ignoring the appropriate coordinate component, instead of actually performing the matrix operation.

Multiview projection is an orthographic projection for which the object is behind the plane of projection, and is oriented so that only two of its dimensions are shown. Generally, three views of an object are drawn, the features and dimensions in each view accurately representing those of the object.

The front view of an object shows the width and height dimensions, as illustrated in Fig. 5.7. The frontal plane of projection is the plane onto which the front view of a multiview drawing is projected.

The top view of an object shows the width and depth dimensions, as illustrated in Fig. 5.8. The top view is projected onto the horizontal plane of projection.

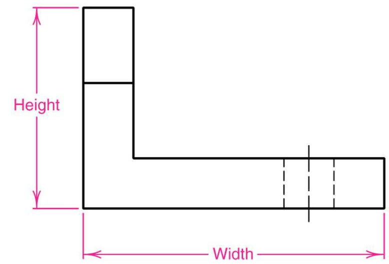


Figure 5.7: Front view (single view)

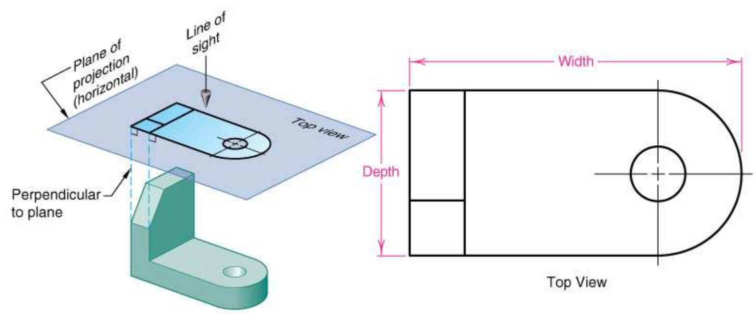


Figure 5.8: Top view

The side view of an object shows the height and depth dimensions, as illustrated in Fig. 5.9. The side view is projected onto the profile plane of projection. The right side view is the standard side view normally used in North America.

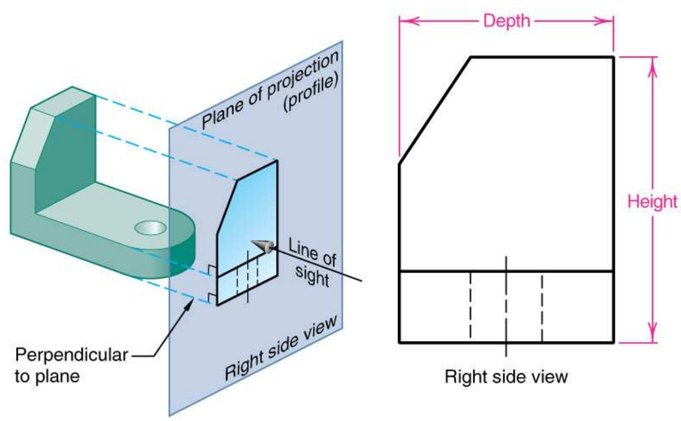


Figure 5.9: Side view

The top view is always positioned above and aligned with the front view, the right side view being always positioned to the right of and also aligned with the front view, as shown in Fig. 5.10, for the same object of Fig. 5.8 and 5.9.

The advantage of multiview drawings over pictorial drawings is that multiview drawings show the true size and shape of the various features of the object, whereas pictorials distort true dimensions, which are critical in manufacturing and construction.

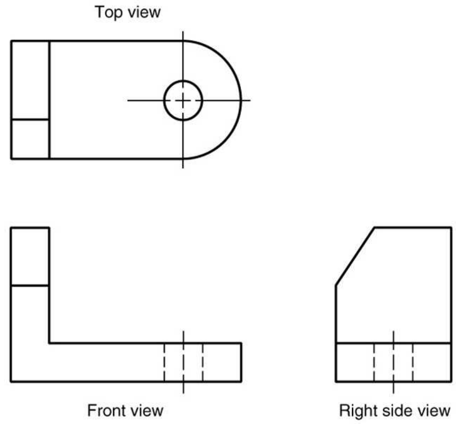


Figure 5.10: Multiview drawing of an object: The North-American way

In Fig. 5.11, we can see an example of a multiview drawing.

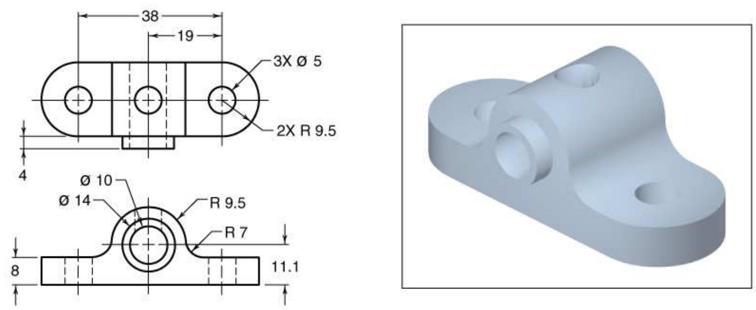


Figure 5.11: Example of multiview drawing

5.2.2 Axonometric Projection

In an axonometric projection, the planes of the object are inclined with respect to the projection plane. Of the axonometric projection types possible, the most commonly used in engineering is the isometric projection. In this type of projection, the angles between the principal axes are all equal to 120°.

To obtain an isometric projection using computational methods, a series of rotations, translations, or both, are performed on the object. A *foreshortening factor* is given by the ratio of the projected length of each line to its true length.

Assuming a projection onto the XY-plane, the necessary rotations can be produced as a θ_y rotation about the Y-axis first, followed by a θ_x rotation about the X-axis, as illustrated in Fig. 5.12. This sequence will maintain the verticality of lines in the projection, a standard technique for showing isometric drawings in engineering. The total operation can be referred to as "tilting", and represented by the homogeneous transformation matrix $\boldsymbol{T}$. The appropriate rotation matrices that cause this tilt are derived below:

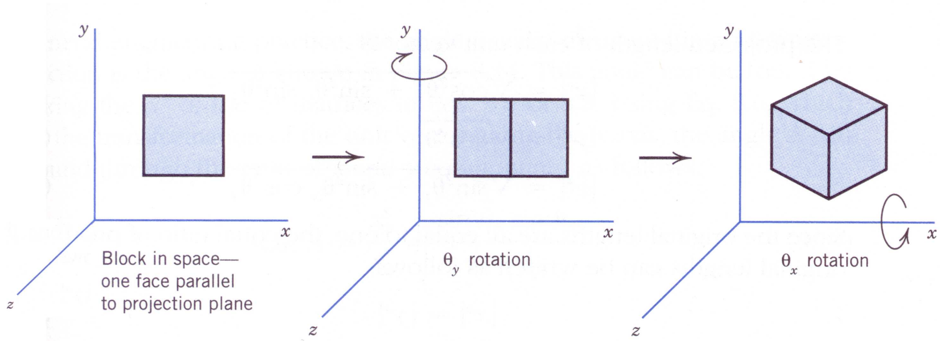


Figure 5.12: Transformations required to obtain an isometric projection

$$
\begin{aligned}
\boldsymbol{T} &= \boldsymbol{Q}_x \boldsymbol{Q}_y \\
&= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x & 0 \\ 0 & \sin\theta_x & \cos\theta_x & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y & 0 \\ \sin\theta_x \sin\theta_y & \cos\theta_x & -\sin\theta_x \cos\theta_y & 0 \\ -\sin\theta_y \cos\theta_x & \sin\theta_x & \cos\theta_x \cos\theta_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\end{aligned}
\tag{5.4}
$$

To finalize the isometric view, an orthographic projection onto the XY plane is obtained by:

$$
\begin{aligned}
\boldsymbol{T}_{ISO} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y & 0 \\ \sin\theta_x \sin\theta_y & \cos\theta_x & -\sin\theta_x \cos\theta_y & 0 \\ -\sin\theta_y \cos\theta_x & \sin\theta_x & \cos\theta_x \cos\theta_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y & 0 \\ \sin\theta_y \sin\theta_x & \cos\theta_x & -\sin\theta_x \cos\theta_y & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\end{aligned}
\tag{5.5}
$$

which is the matrix representing the isometric projection.

5.2.3 Oblique Projection

Both the multiview orthographic and the axonometric projections are created with the projectors perpendicular to the plane of projection. An *oblique projection* has the projectors at an angle with the plane of projection, as shown in Fig. 5.13.

The general formulation for the oblique projection will be derived by considering the point $P(0,\ 0,\ 1)$ with XY being the *projection plane*.

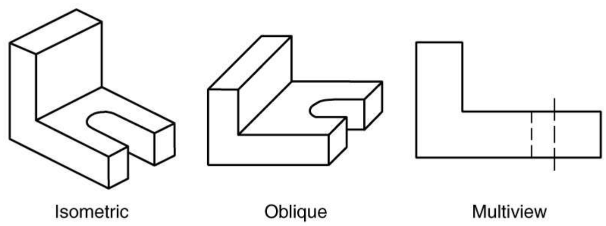


Figure 5.13: Parallel projection techniques

With reference to Fig. 5.14, the distance l gives the *foreshortening ratio* of any line perpendicular to the $z = 0$ plane, after projection. If θ is the angle between the projection $\overline{OP^*}$ of segment $\overline{OP}$ and the X-axis, then the *oblique-projection matrix* $\boldsymbol{T}_{OBL}$ becomes

$$\boldsymbol{T}_{OBL} = \begin{bmatrix} 1 & 0 & l\cos\theta & 0 \\ 0 & 1 & l\sin\theta & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.6}$$

and the oblique projection of P is, in homogeneous coordinates,

$$\boldsymbol{p}^* = \boldsymbol{T}_{OBL} \begin{bmatrix} l\boldsymbol{k} \\ 0 \end{bmatrix} = \begin{bmatrix} x* \\ y* \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} l\cos\theta \\ l\sin\theta \\ 0 \\ 0 \end{bmatrix} \tag{5.7}$$

with $\boldsymbol{k}$ defined as in eq. (1.7).

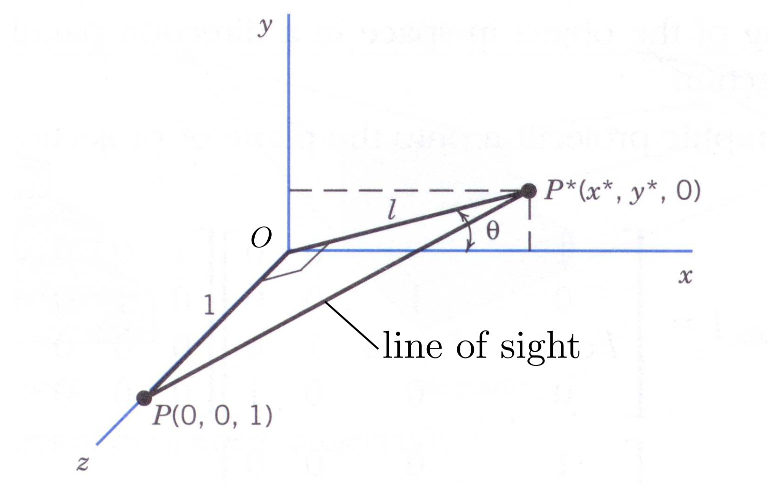


Figure 5.14: Oblique projection of a point P on the XY-plane

If l, the foreshortening ratio, is equal to 1, lines perpendicular to the projection plane preserve their original length: this is the *cavalier projection*. If $l = 1/2$, the projection length of lines perpendicular to the projection plane is half their original length: this is the *cabinet projection*.

Note that the value of θ is independent of l. The most commonly used values for θ lie between 30° and 45°.

It is noteworthy that the oblique projection matrix is also a shearing matrix. In fact, application of the oblique projection matrix causes shearing of the object in space.

Summary

Parallel projections have traditionally been used in engineering practice. In some cases, they preserve the true dimensions of an object but do not produce a realistic picture. The perspective projection, not included in these notes, gives the exact opposite effect: realistic image but loss of true dimensions.

5.3 Visualization

5.3.1 The Six Principal Views

There are six principal mutually perpendicular views projected onto three mutually perpendicular projection planes. These views are the *top*, *front*, *right*, *left*, *bottom* and *back*, as depicted in Fig. 5.15.

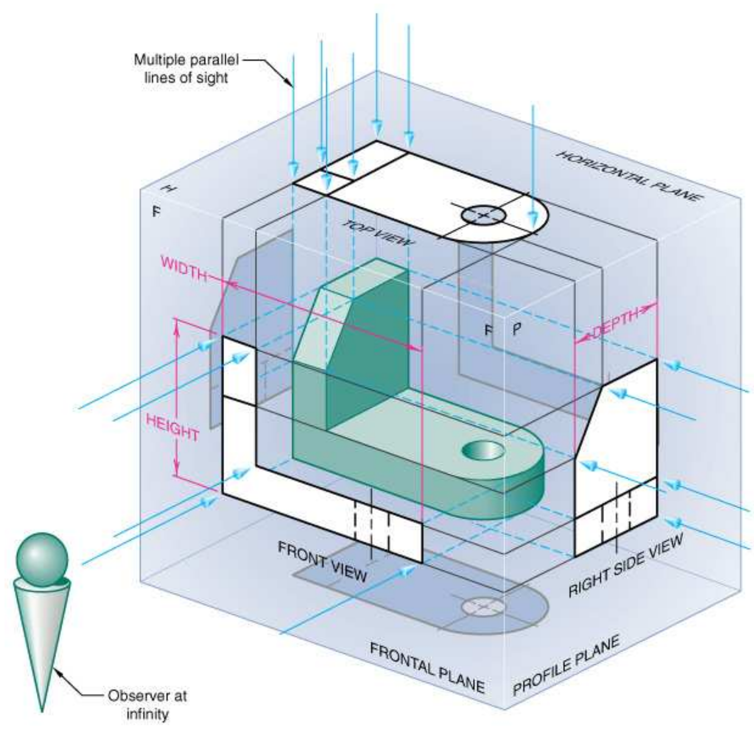


Figure 5.15: Object producing the six principal views

The width dimension is common to the front and top views. The height dimension is common to the front and side views. The depth dimension is common to the top and side views, as shown in the six-view drawing of Fig. 5.16.

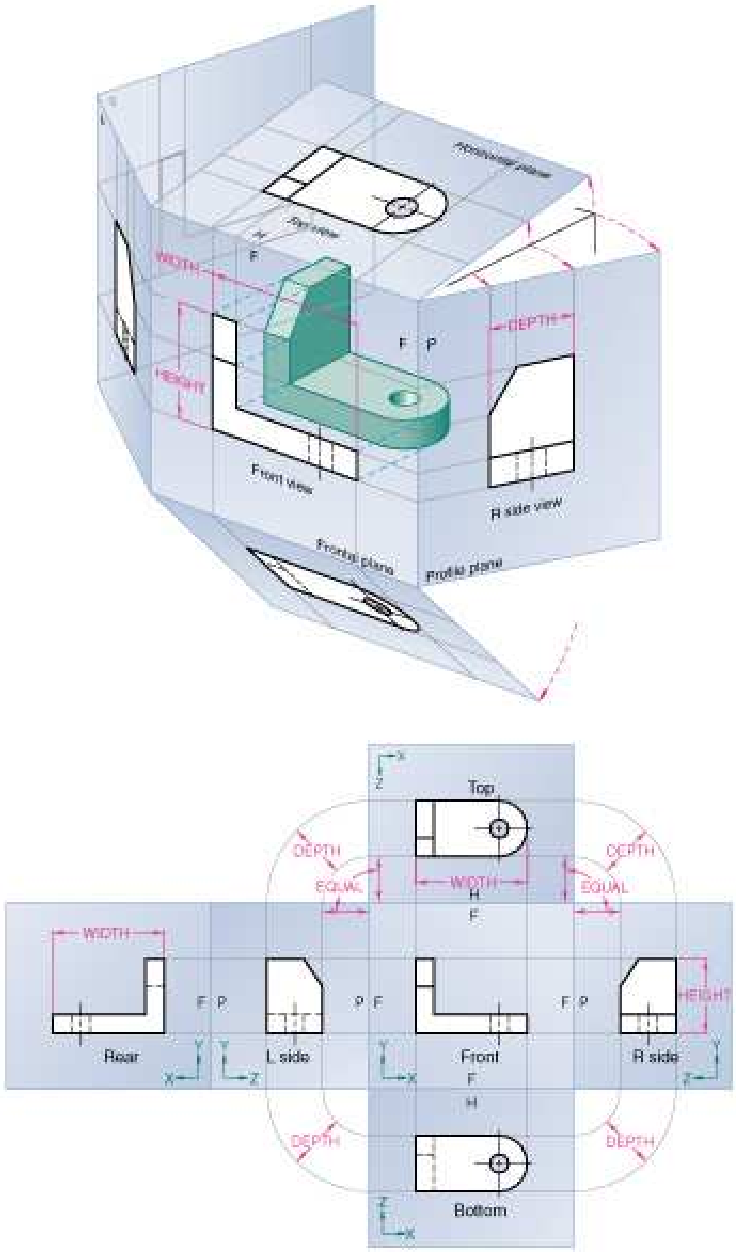


Figure 5.16: The six-view drawing

The arrangement of views may vary as long as the dimension-alignment is correct, as shown in Fig. 5.17. The organization of the views may vary as well; for example, you can look at the alternative view arrangement in Fig. 5.18.

Third-angle projection is the standard projection for the United States and Canada. The ANSI third angle icon is shown in Fig. 5.19. *First-angle projection* is the standard in Europe

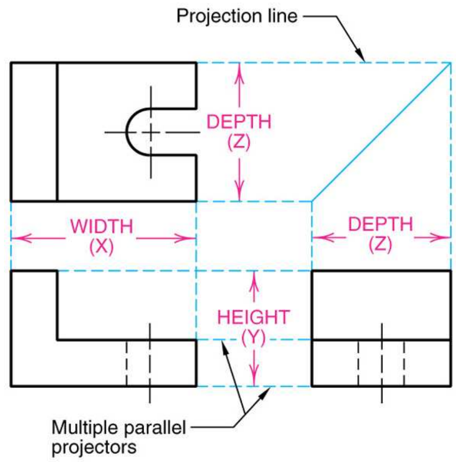


Figure 5.17: Three space dimensions

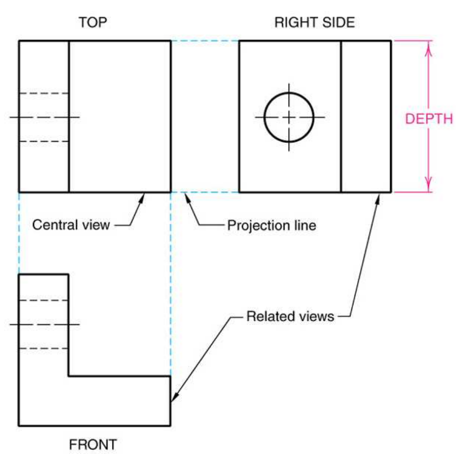


Figure 5.18: Alternate view arrangement

and Asia. The first-angle and third-angle icons are shown in the same figure. The difference between first- and third-angle projections is the placement of the object and the projection plane, as shown in Fig. 5.19.

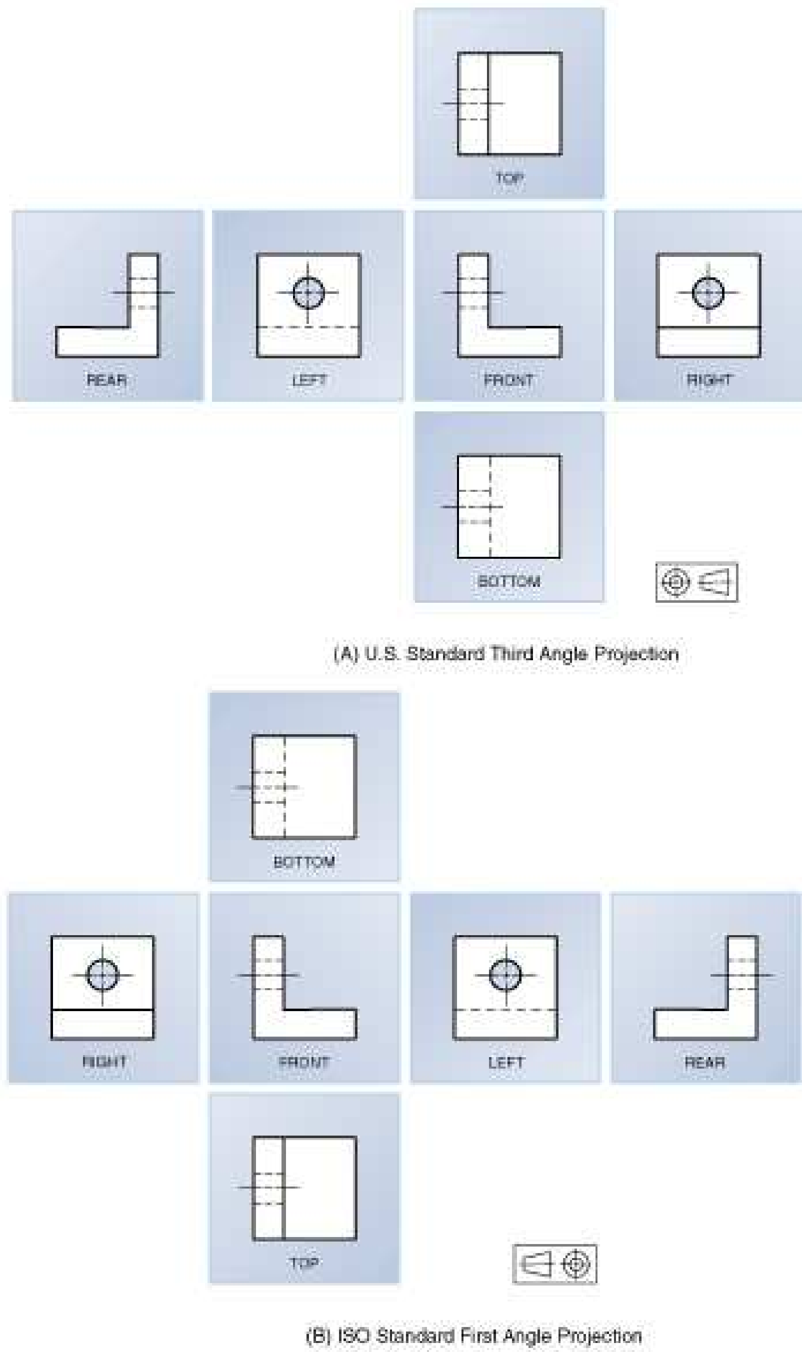


Figure 5.19: Standard arrangement of the six principal views

Adjacent views are two orthographic views placed next to each other such that the dimension they share in common is aligned. Every point or feature in one view must be aligned on a parallel projector in any adjacent view. Related views are two views that share the same adjacent views. Distances between any two points of a feature in related views must be equal. The view from which adjacent views are aligned is the central view, as shown in Fig. 5.20.

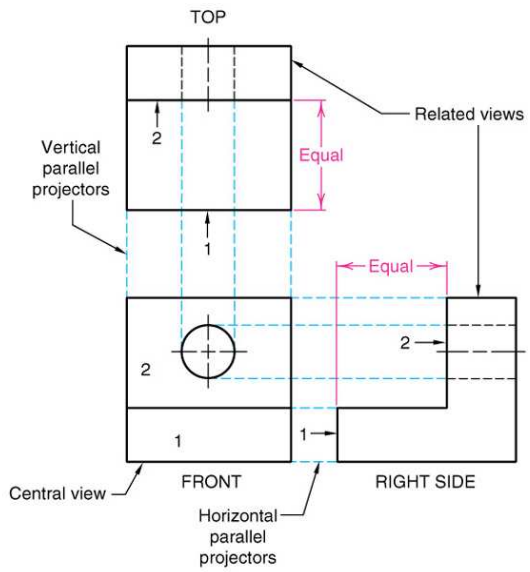


Figure 5.20: Alignment of views

5.3.2 Fundamental views of edges and planes for visualization

- An edge is the intersection of two planes; it is represented as a line segment on a multi-view drawing. A *normal edge*, or *true-length line*, is an edge that is parallel to two planes of projection. An *inclined edge* is parallel to one single plane of projection, but inclined to the adjacent planes and appears foreshortened in the adjacent views. Features are foreshortened when the lines of sight are not perpendicular to the feature. An *oblique edge* is not parallel to any principal plane of projection; therefore, it never appears either as a point or in true-length in any of the six principal views, as depicted in Fig. 5.21.

- A *principal plane* is parallel to one of the principal planes of projection and is therefore perpendicular to the corresponding line of sight. A feature—intrusion, protrusion,bore—lying in a principal plane will be true size and shape in the view where it is parallel to the projection plane; it will appear as a horizontal or vertical line in the adjacent views, as illustrated in Fig. 5.22.

  Since the foregoing feature appears in true size and shape in the front view, it is sometimes referred to as a *normal plane*. This feature also appears as a horizontal edge in the top view and as a vertical edge in the right side view, as shown in Fig. 5.23.

  This edge representation is an important characteristic in multiview drawings, as illustrated in Fig. 5.24.

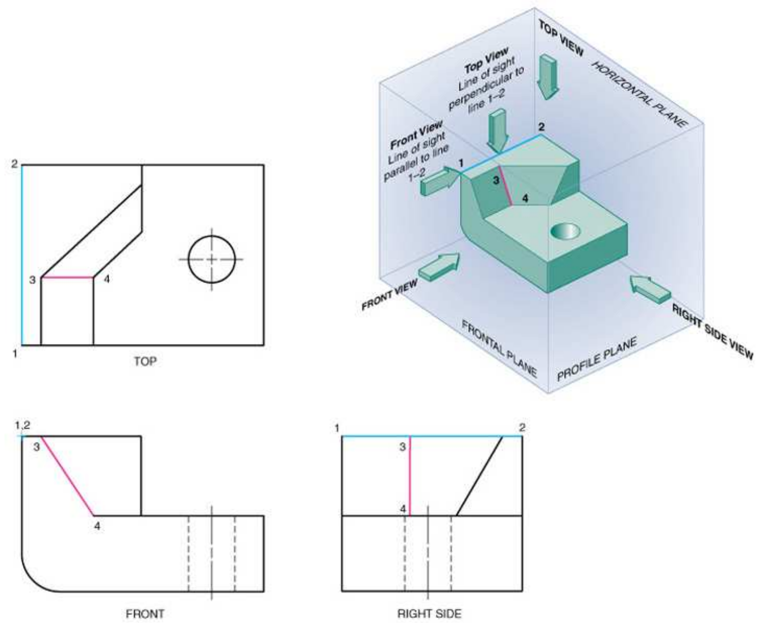


Figure 5.21: Fundamental views of edges

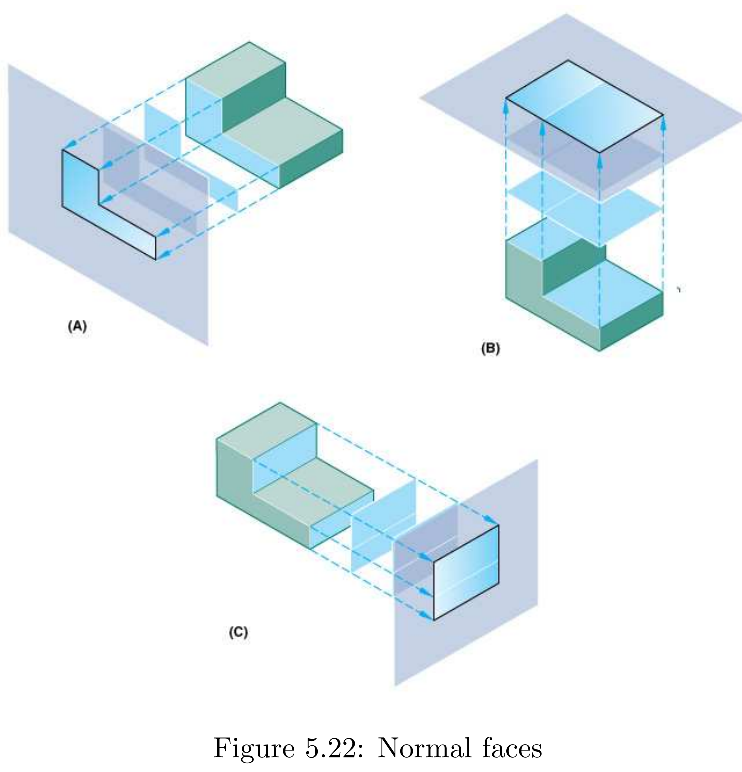


Figure 5.22: Normal faces

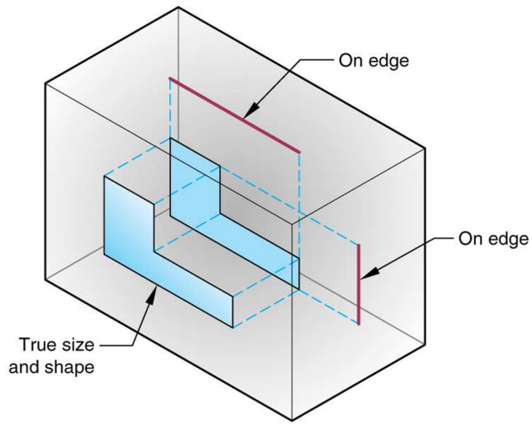


Figure 5.23: Normal face projection

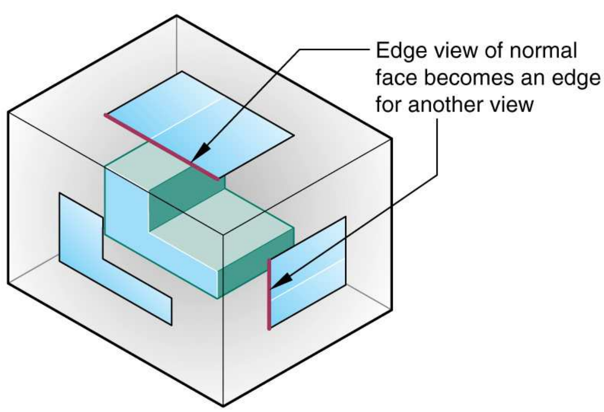


Figure 5.24: Edge views of normal face

Principal planes are categorized by the view in which the planes appear true size and shape: *frontal*, *horizontal*, or *profile*, as shown in Fig. 5.25.

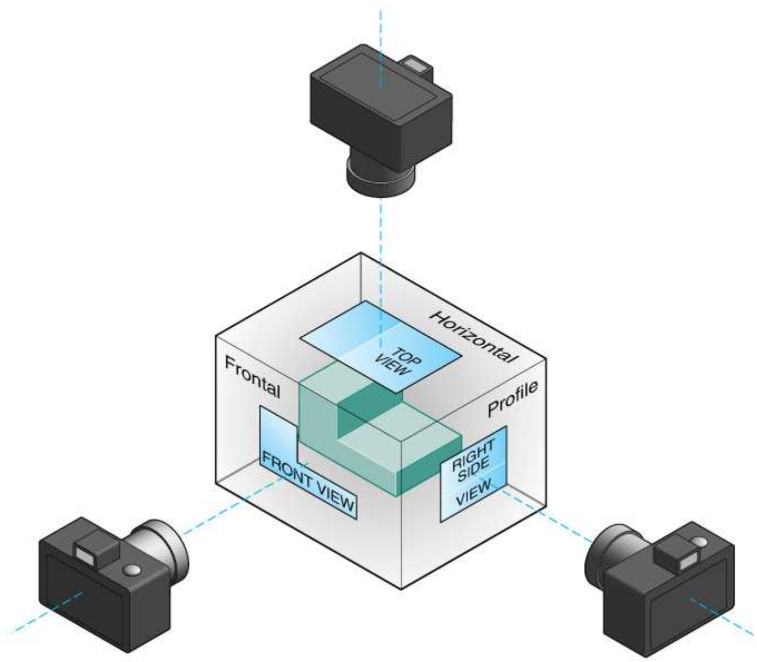


Figure 5.25: The camera metaphor for planes of visualization

Of course, most objects will contain a combination of *principal* (*normal*), *inclined*, and *oblique surfaces*. Figure 5.26 shows how these types of surfaces will be represented in the different views.

5.3.3 Multiview Representations

Three-dimensional solid objects are represented on 2D media by means of points, edges and planes. The solid geometric primitives are transformed into 2D geometric primitives. Figure 5.27 shows multiview drawings of common geometric solids.

- A point represents a specific position in space and has no width, height, or depth. A point can represent:
  - The end view of a line.
  - The intersection of two lines.
  - A specific position in space.
- A plane surface always appears as an edge or a surface. Parallel edges appear parallel in all views. Planes that are parallel to the lines of sight appear as edges. These concepts are shown in Fig. 5.28.

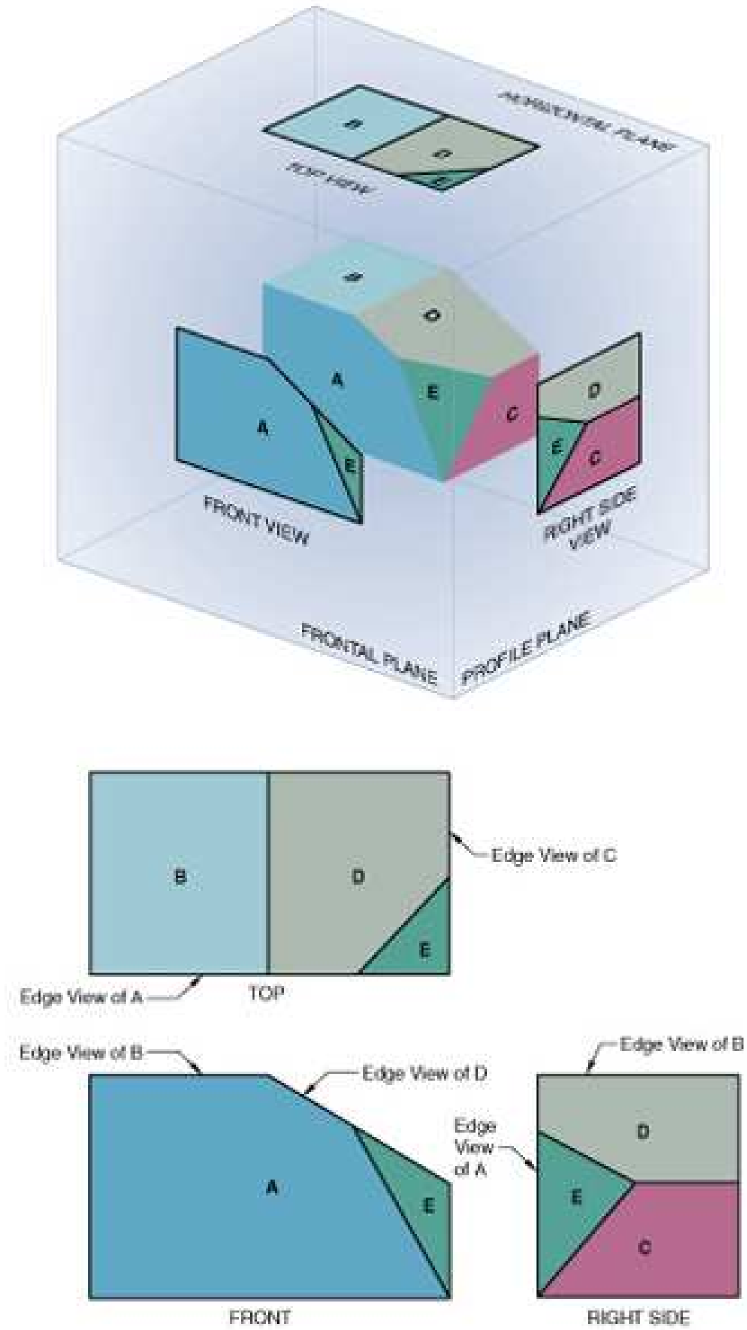


Figure 5.26: Fundamental views of surfaces

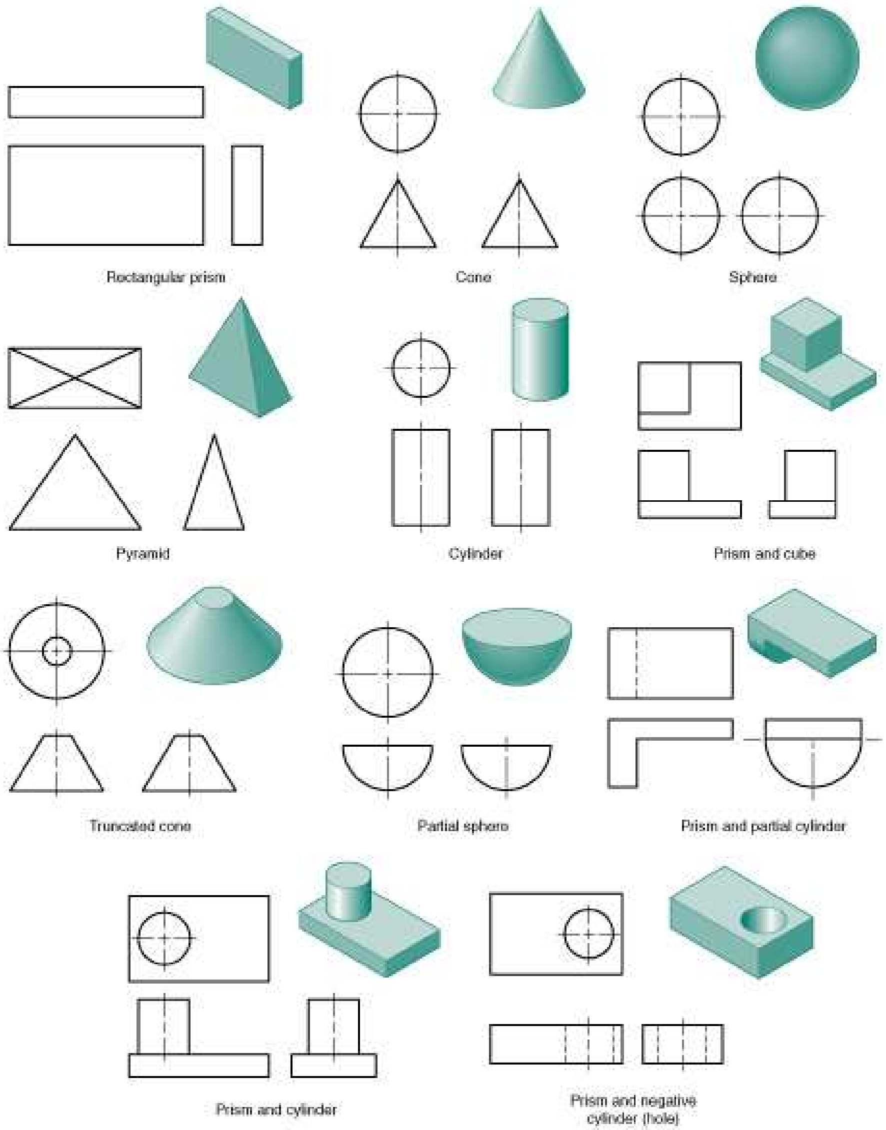


Figure 5.27: Multiview drawings of solid primitive shapes

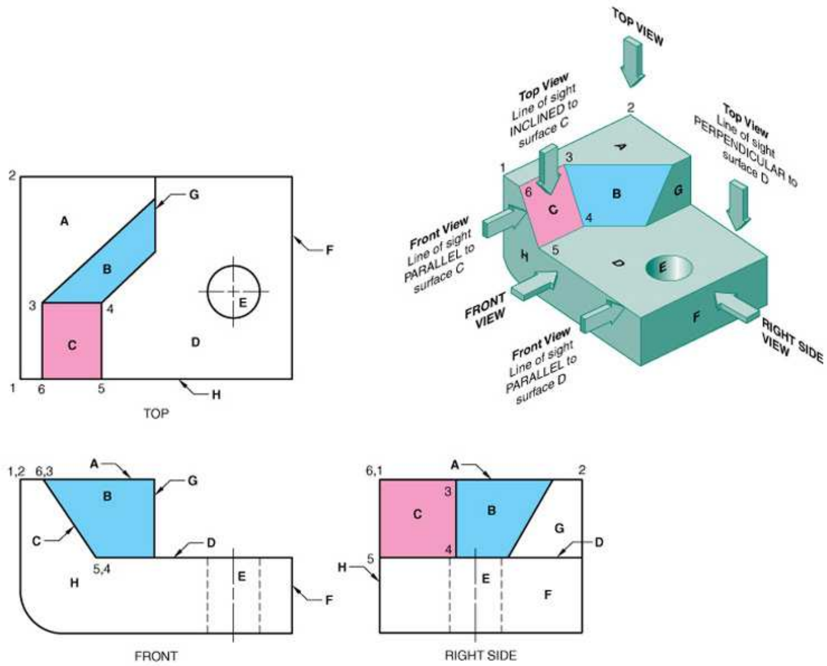


Figure 5.28: Rule of configuration of planes

- Angles are true size when they lie in a normal plane. The rules of angle representation can be observed in Fig. 5.29.

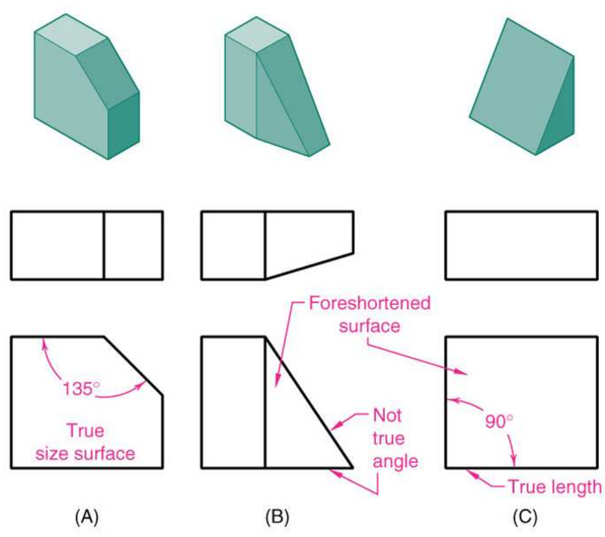


Figure 5.29: Rule of angle representation

- Curved surfaces are used to show drilled holes and cylindrical features. Only the far outside boundary—the limiting element—of a curved surface is represented in multiview drawings, as illustrated in Fig. 5.30.

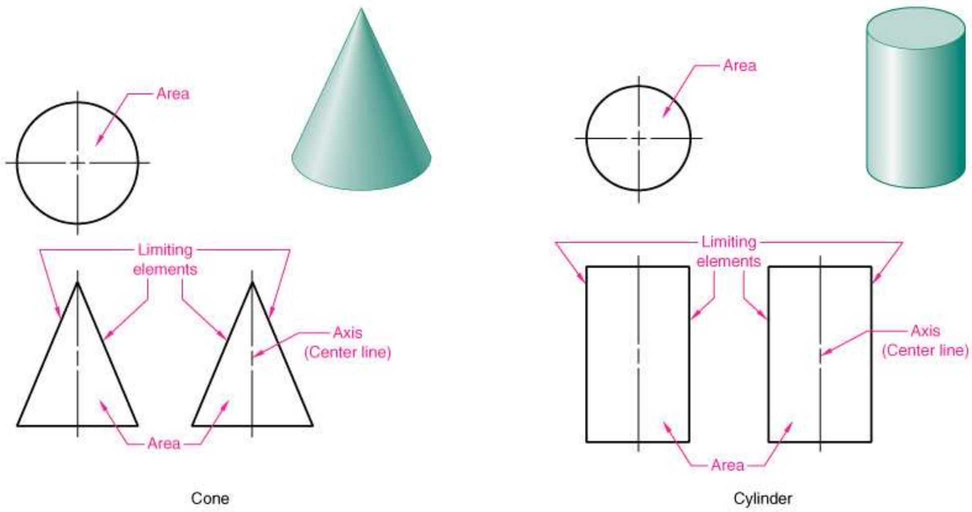


Figure 5.30: Limiting elements

- A partial cylinder with its bases lying on a principal plane is represented by an arc and an edge on the corresponding plane of projection; it is represented by rectangles in the adjacent views. *If the partial cylinder is tangent to a neighbouring surface, then no edge is shown along the tangent*, as in Fig. 5.31; however, if tangency does not exist, then an edge is used to represent the abrupt change of surface orientation between the partial cylinder and the neighbouring surface, as shown in Fig. **??**.

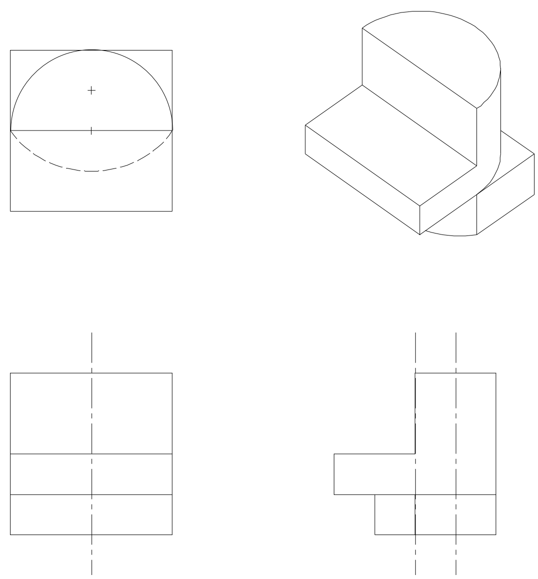


Figure 5.31: Tangent partial cylinder

- An ellipse is used to represent a hole or circular feature that is viewed at an angle other than perpendicular or parallel. For example, Fig. 5.32 shows the end of a cylinder, viewed first with a perpendicular line of sight and then at 45°. From the perpendicular view, the centre lines are true length, and the figure is represented as a circle, as in Fig. 5.33.

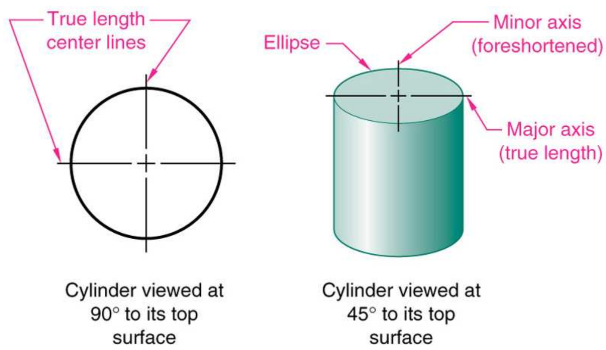


Figure 5.32: Elliptical representation of a circle

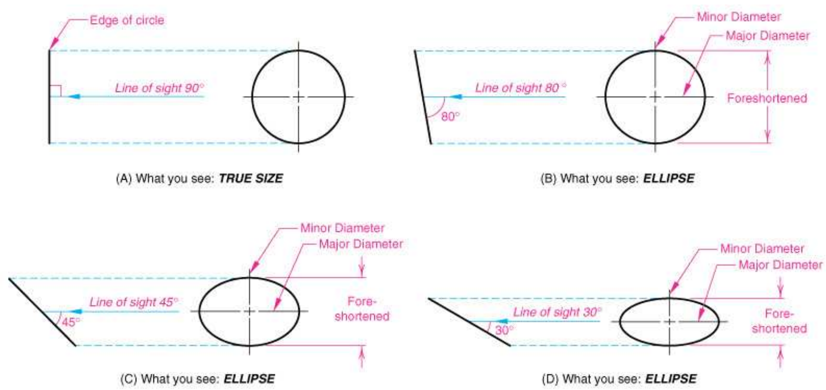


Figure 5.33: Viewing angles for ellipses

- Holes follow standards and conventions of representation that are represented in Fig.5.34:
  - A **through hole** is a hole that goes all the way through an object, is represented in one view as two parallel hidden lines for the limiting elements, and is shown as a circle in the adjacent view.
  - A **blind hole** is a hole that is not drilled all the way through the object.
  - **Counterbored** holes are used to allow the heads of bolts to be flush or below the surface of the part.
  - **Countersunk** holes are commonly used for flathead screws, and are represented by 45° lines.
  - A **spotface** hole provides a place for heads of fasteners to rest by creating a smooth surface on cast parts.
  - A **threaded hole** is represented with 2 hidden lines in the front view, and a solid and a hidden line in the top view.

238 CHAPTER 5

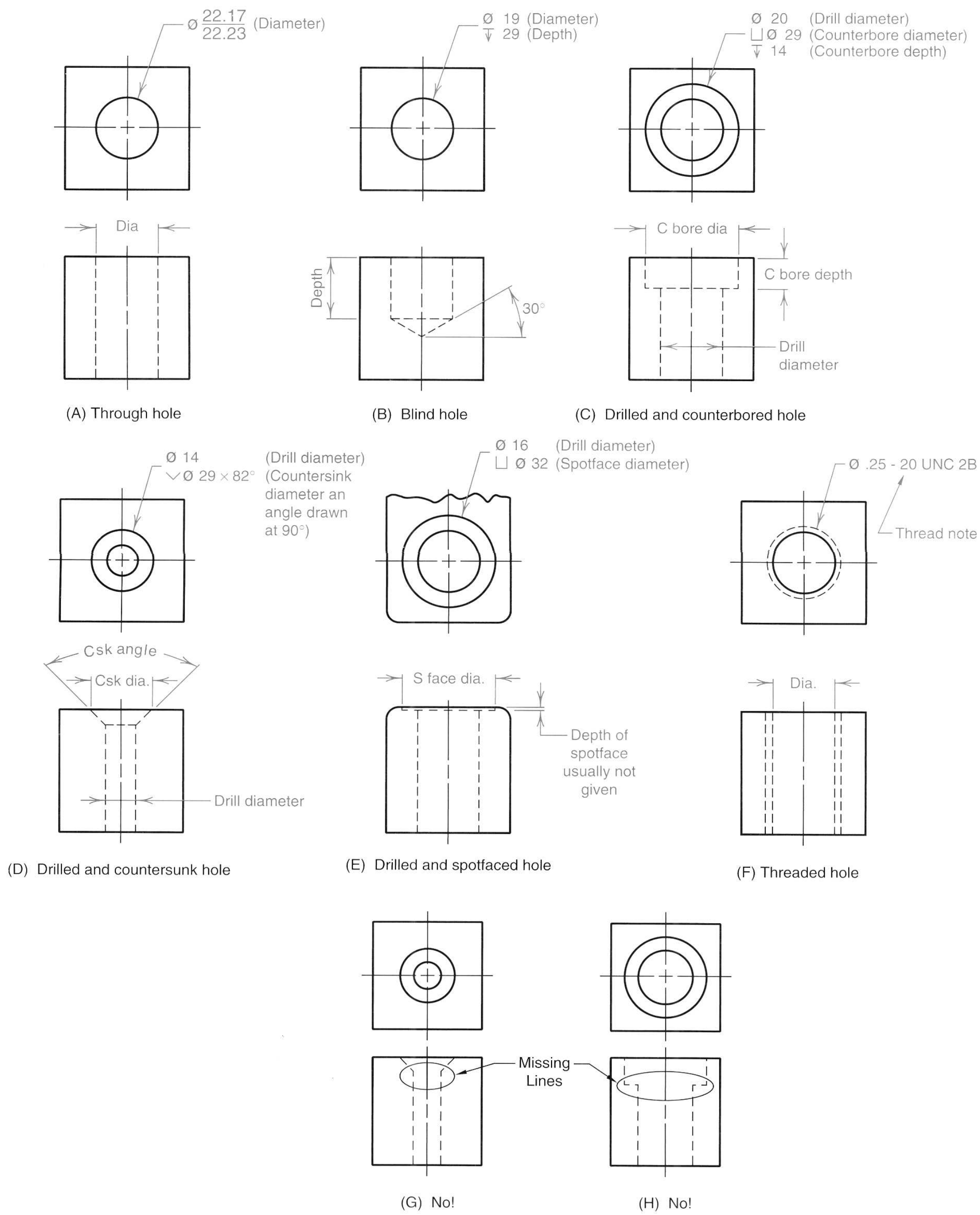


Figure 5.61 Representation of various types of machined holes

Figure 5.34: Representation of various types of machined holes

In all hole representations a line must be drawn to represent the change that occurs between the large and small diameter.

Index