*Matric notation is used to develop geometric concepts for computer-controlled graphics.*

*This natural form of geometric expression leads to homogeneous coordinates which form the basis of an algorithm used for geometric construction. In this way, three-dimensional objects and other pictures can be displayed on a graphics console.*

*The paper also briefly discusses the notation and development of functions for the construction of surfaces.*

**INTERACTIVE GRAPHICS IN DATA PROCESSING**

# Geometry for construction and display

**by D. V. Ahuja and S. A. Coons**

The design and delineation of arbitrary shapes, and the display of these shapes from arbitrarily chosen viewpoints are more easily accomplished if done on the graphic display console of a computer rather than by pencil and paper methods. In this paper, we show modern methods for geometry, so constructed and organized that geometric manipulations can be performed in a way natural to the computer, and can yield results that are natural to man.

We discuss three major ideas: matric methods for algebra, homogeneous coordinates, and curves and surfaces from a parametric standpoint. First, we introduce matric methods for algebra, because matrices exhibit in a very transparent way the nature of geometric entities. They are also natural forms for computer implementation, since matrix manipulations are already programmed, and routines exist that are easily exploited.

Matrices lead naturally to the notion of coordinate transformations, and this notion leads into homogeneous coordinates. On this basis, we are able to show that all possible pictures, or displays, of three-dimensional objects, including the projections employed by draftsmen, so-called "axonometric pictorials," and pictures in full perspective can be constructed on the display console of the computer by a single, simple algorithm.

Finally, we briefly discuss curves and surfaces from a parametric standpoint. This form of curve and surface equations is used in "differential" geometry, and it lends itself to computer implementation in a very natural way.

## Matric notation

We begin with a very elementary discussion of matric notation. In the plane, a line can be represented by the equation:

$$Ax + By = C$$

In matric notation, this equation becomes

$$[x \ \ y]\begin{bmatrix} A \\ B \end{bmatrix} = C$$

The brackets enclose two matrices, and their entries are the quantities that appear in the ordinary algebraic equation. The bracketed matrices are juxtaposed, which suggests that they are to be multiplied together; the "product" of the matrices is just $Ax + By$.

Now assume there are *two* lines in the plane, given by the linear equations:

$$Ax + By = C$$
$$Dx + Ey = F$$

In matric notation, these equations become the single matric equation

$$[x \ \ y]\begin{bmatrix} A & D \\ B & E \end{bmatrix} = [C \quad F]$$

Ordinarily, there exists a point of intersection of these lines. This is to say that provided the lines are not parallel (we dispose of this case later, and show that it is not exceptional), it is possible, for $A\ B\ D\ E\ C\ F$ given numbers, to choose a pair of numbers $x$ and $y$ that "satisfy" both equations simultaneously. This algebraic process is very elementary and well-known, but it is interesting to see how it is done with matrices.

We can illustrate this process by an example. We wish to find the intersection of the lines:

$$x + y = 1$$
$$y = \tfrac{2}{3}x$$

The last equation can be written as

$$2x - 3y = 0$$

In matric form, we have

$$[x \ \ y]\begin{bmatrix} 1 & 2 \\ 1 & -3 \end{bmatrix} = [1 \quad 0]$$

If we multiply both sides of this equation by the *inverse* of matrix $\begin{bmatrix} 1 & 2 \\ 1 & -3 \end{bmatrix}$ we get without destroying its validity matrix operators

$$[x \ \ y]\begin{bmatrix} 1 & 2 \\ 1 & -3 \end{bmatrix}\begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & -\frac{1}{5} \end{bmatrix} = [1 \quad 0]\begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & -\frac{1}{5} \end{bmatrix}$$

or

$$[x \quad y]\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [\tfrac{3}{5} \quad \tfrac{2}{5}]$$

The matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is called the *identity matrix*. Any matrix multiplied by the identity matrix is unchanged. Hence, $[x \quad y] = [\tfrac{3}{5} \quad \tfrac{2}{5}]$, which is the point of intersection of these lines. However, not all matrices have inverses. For instance, the matrix $\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$ has no inverse. It is called a *singular matrix*. Its determinant is equal to zero, and indeed, the vanishing of the determinant of a matrix is a necessary and sufficient condition to test singularity. But even when matrices do not have inverses, we can obtain solutions to equations.

The equation for a line,

$$[x \quad y]\begin{bmatrix} A \\ B \end{bmatrix} = C$$

can be written

$$[x \quad y \quad 1]\begin{bmatrix} A \\ B \\ -C \end{bmatrix} = 0$$

without destroying the equality. When we multiply the matrices, we obtain

$$Ax + By - C = 0$$

Now forgetting the sign of $C$, consider the expression

$$[x \quad y \quad 1]\begin{bmatrix} A \\ B \\ C \end{bmatrix} = 0$$

**coordinates**

The matrix $[x \quad y \quad 1]$ consists of the coordinates of points that lie on a line; any pair of numbers $x$ and $y$ that satisfy this equation lie on this line. For this reason, we call $[x \quad y \quad 1]$ point coordinates. But suppose we were to fix $x$ and $y$; then any set of numbers $A\ B\ C$, properly chosen, would represent lines passing through the fixed $x$-$y$ point. For this reason, $[A \quad B \quad C]$ are called *line coordinates*. There are only two coordinates for the point, $x$ and $y$, but there are apparently three coordinates for the line. We are already aware of the two statements, "two lines determine a point" and "two points determine a line." These statements are in a sense symmetric (in geometry, they are known as "dual" statements), and it would be more satisfying if we could observe a similar symmetry in the equation. We might write:

$$[x \quad y \quad 1]\begin{bmatrix} A/C \\ B/C \\ 1 \end{bmatrix} = 0 \qquad \text{or} \qquad [x \quad y \quad 1]\begin{bmatrix} P \\ Q \\ 1 \end{bmatrix} = 0$$

This representation would be acceptable, except that it would not permit us to write an equation for a line passing through the origin. Such a line is:

$$Ax + By = 0 \qquad \text{or} \qquad [x \quad y \quad 1]\begin{bmatrix} A \\ B \\ 0 \end{bmatrix} = 0$$

Since we need $A\ B\ C$, three coordinates for the line, it appears that perhaps we really need three coordinates for the point as well:

$$[x \quad y \quad w]\begin{bmatrix} A \\ B \\ C \end{bmatrix} = 0$$

For ordinary points we can set $w = 1$, and for lines not through the origin, we can set $C = 1$ also. When $C = 0$, we have an equation for lines through the origin. When $w = 1$, $x$ and $y$ are the *ordinary* coordinates of points in the plane. When $w \neq 1$, we can obtain the ordinary coordinates by putting $[x \quad y \quad w]$ into the form $[x_o \quad y_o \quad 1]$ where $x_o$, $y_o$ are the ordinary coordinates. Obviously we can do this without destroying the validity of the equation by multiplying by $1/w$ provided $w$ is not zero, i.e., if

$$[x \quad y \quad w]\begin{bmatrix} A \\ B \\ C \end{bmatrix} = 0$$

then

$$\frac{1}{w}[x \quad y \quad w]\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} \frac{x}{w} & \frac{y}{w} & 1 \end{bmatrix}\begin{bmatrix} A \\ B \\ C \end{bmatrix} = 0$$

Then $x/w = x_o$ and $y/w = y_o$, the sought-for ordinary coordinates. Since the ordinary coordinates are obtained by $x_o = x/w$ and $y_o = y/w$, we modify the notation slightly. We can thus write $[wx \quad wy \quad w]$ instead of $[x \quad y \quad w]$.

In this way, we are able to keep track of the ordinary coordinates of a point; we consider $wx$ and $wy$ as biliteral symbols throughout all calculations and perform the division $wx/w$ and $wy/w$ only at the end. With this notation, we can retain $x$, $y$ as ordinary coordinates instead of $x_o$, $y_o$. But if $w = 0$, we cannot perform the division. What does the set of numbers $[wx \quad wy \quad w]$ then mean?

To experiment with this question, let us try a specific example. Let $wx$ and $wy$ be 2 and 3, but let $w$ vary, and calculate the ordinary $x$ and $y$ coordinates. The results appear in Table 1.

From this table, it is clear that as $w$ decreases to 0, $x$ and $y$ increase without bound, but always in the ratio 2 : 3. Evidently

**Table 1** x and y coordinates

| $w$ | $x$ | $y$ |
|---|---|---|
| 1 | 2 | 3 |
| $\frac{1}{2}$ | 4 | 6 |
| $\frac{1}{3}$ | 6 | 9 |
| . | . | |
| . | . | |
| . | . | |
| $\frac{1}{10}$ | 20 | 30 |
| . | | |
| . | | |
| . | | |
| $\frac{1}{100}$ | 200 | 300 |

$[2 \quad 3 \quad 0]$ represents a point at infinity. The ordinary coordinates of such a point are of no use to us, because they are $[\infty \quad \infty \quad 1]$, and the infinity signs cannot be manipulated algebraically.

### Homogeneous coordinates

The set of numbers $[2 \quad 3 \quad 0]$ is a matrix of coordinates for a point at infinity that is not algebraically exceptional, as is shown later. The set of coordinates

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix}$$

and the set of coordinates $[wx \quad wy \quad w]$ are called the *homogeneous* coordinates of line and point, respectively. Homogeneous refers to the algebraic forms that they constitute; all entries are of the same dimension. Indeed, it is appropriate to think of $[wx \quad wy \quad w]$ as a three-dimensional matrix.

We return to the two lines in the plane, which give the homogeneous equations:

$$[wx \quad wy \quad w]\begin{bmatrix} 1 & 2 \\ 1 & -3 \\ -1 & 0 \end{bmatrix} = [0 \quad 0]$$

We cannot solve this equation for $wx$ and $wy$ by finding the inverse of the $3 \times 2$ matrix because it is not a square matrix and has no ordinary inverse. However, we can arrange it so that the matrix does have an inverse by adjoining a column to it:

$$[wx \quad wy \quad w]\begin{bmatrix} 1 & 2 & 0 \\ 1 & -3 & 0 \\ -1 & 0 & 1 \end{bmatrix} = [0 \quad 0 \quad w]$$

This step is harmless and does not violate the equality. The inverse of this new square matrix can be found and is the matrix

$$\tfrac{1}{5}\begin{bmatrix} 3 & 2 & 0 \\ 1 & -1 & 0 \\ 3 & 2 & 5 \end{bmatrix}$$

Then

$$[wx \quad wy \quad w] = \tfrac{1}{5}[0 \quad 0 \quad w]\begin{bmatrix} 3 & 2 & 0 \\ 1 & -1 & 0 \\ 3 & 2 & 5 \end{bmatrix} = \tfrac{1}{5}[3w \quad 2w \quad 5w]$$

$$= \begin{bmatrix} \frac{3w}{5} & \frac{2w}{5} & w \end{bmatrix}$$

Dividing both sides by $w$ to obtain the ordinary coordinates of the point of intersection, we have $[x \quad y \quad 1] = [\frac{3}{5} \quad \frac{2}{5} \quad 1]$ as before.

**parallel lines**

Consider the following two parallel lines. (In ordinary algebraic notation, they are: $x + y = 1$ and $y = -x$.)

$$[wx \quad wy \quad w]\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ -1 & 0 \end{bmatrix} = [0 \quad 0]$$

We can now proceed to determine their point of intersection. We prepare the matrix to read

$$[wx \quad wy \quad w]\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} = [0 \quad 0 \quad w]$$

The matrix on the left is still singular, and we cannot invert it. (It is singular because two rows are identical, and its determinant still vanishes.) We change the modification of the matrix, so that it reads

$$[wx \quad wy \quad w]\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} = [0 \quad 0 \quad wx]$$

Now one of the unknowns, $wx$, appears on the right side of the equation.

The matrix has an inverse, and we obtain the solution:

$$[wx \quad wy \quad w] = [0 \quad 0 \quad wx]\begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} = [wx \quad -wx \quad 0]$$

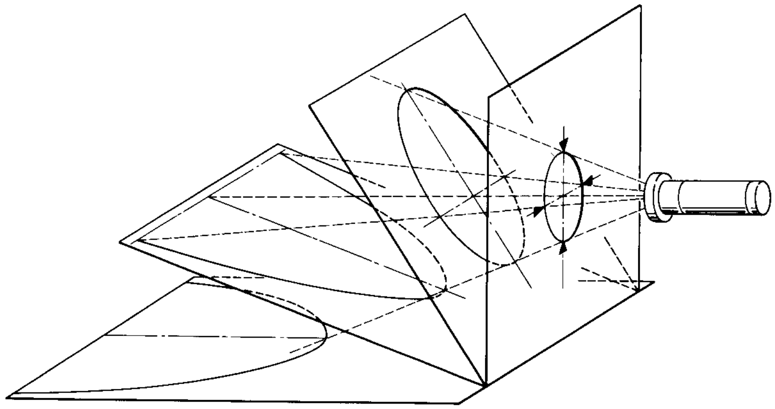
Consider the point represented by the matrix $[wx \quad -wx \quad 0]$. This matrix is the same as $wx\,[1 \quad -1 \quad 0]$, where $wx$ is any number whatever. It is, as we have seen, a point at infinity.

The preceding elementary discussion indicates that homogeneous coordinates are useful as well as aesthetically satisfying. They make equations more symmetric in form, which appeals to our sense of structure and completeness; they permit us to deal with infinitely distant points in the plane as easily as we deal with local points; and they permit us to remove the exceptional cases. In the following discussion of projective transformations, we see that they have even greater value.

**perspective transformation**

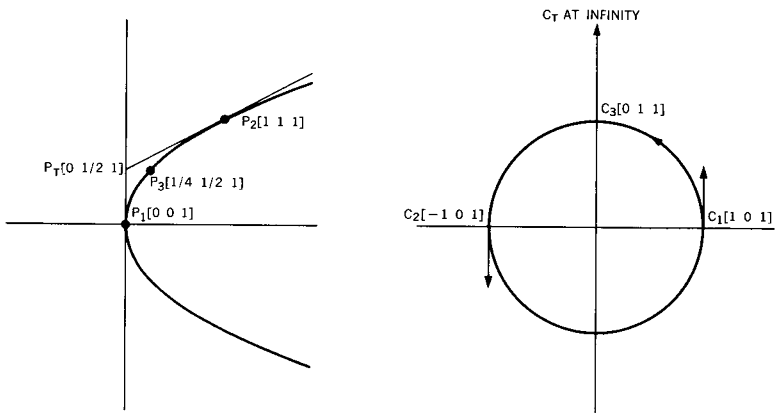
Let us first consider the perspective projections of a plane on to a plane by taking an example. From the theory of conic sections, we know that the circle, ellipse, hyperbola, and parabola are perspective projections of one another as shown in Figure 1.[1] We accomplish this transformation mathematically using homogeneous coordinates and the matric notation.

**Figure 1** Perspective projections of a circle*



*USED BY PERMISSION OF McGRAW-HILL BOOK COMPANY

**Figure 2** Perspective transformation




For simplicity, let us consider a parabola $x = y^2$, shown in Figure 2, as our starting curve, and transform this parabola into any other conic section. The parabola $x = y^2$ may be expressed parametrically as $x = u^2$ where $y = u$.

Any point $[x \quad y \quad 1]$ on this parabola is given in matric notation by:

$$[x \quad y \quad 1] = [u^2 \quad u \quad 1]\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = [u^2 \quad u \quad 1]$$

We transform this base parabola into any other conic section (including another parabola) by transforming four arbitrarily chosen points from the plane "$R$" of the parabola into four points in another plane "$C$" of any conic section. Let us choose points $P_1$ $(u = 0)$, $P_2$ $(u = 1)$, $P_3$ $(u = \frac{1}{2})$, and $P_T$ (the point of intersection of tangents at $P_1$ and $P_2$) which transform into corresponding points $C_1$, $C_2$, $C_3$, and $C_T$ in the plane $C$. We use the previously established homogeneous coordinate notation for the general point in plane $C$ — for example, $C_1 = [w_1x_1 \quad w_1y_1 \quad w_1]$ which corresponds to the ordinary point $[x_1 \quad y_1 \quad 1]$.

We state that there exists a matrix $A$ that accomplishes the previously described transformation as follows:

$$PA = C$$

or

$$[u^2 \quad u \quad 1]A = [wx \quad wy \quad w]$$

Evidently $A$ is a $3 \times 3$ matrix. Let us proceed to evaluate $A$. For the transformation of $P_1$ into $C_1$, the above equation is

$$[0 \quad 0 \quad 1]A = [w_1x_1 \quad w_1y_1 \quad w_1] \tag{1}$$

Similarly for points $P_2$ and $P_T$

$$[1 \quad 1 \quad 1]A = [w_2x_2 \quad w_2y_2 \quad w_2] \tag{2}$$

$$[0 \quad \tfrac{1}{2} \quad 1]A = [w_Tx_T \quad w_Ty_T \quad w_T] \tag{3}$$

We can combine Equations 1, 2, and 3 and write

$$\underset{B}{\begin{bmatrix} 0 & 0 & 1 \\ 0 & \frac{1}{2} & 1 \\ 1 & 1 & 1 \end{bmatrix}} A = \begin{bmatrix} w_1x_1 & w_1y_1 & w_1 \\ w_Tx_T & w_Ty_T & w_T \\ w_2x_2 & w_2y_2 & w_2 \end{bmatrix}$$

$$= \begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_T & 0 \\ 0 & 0 & w_2 \end{bmatrix} \begin{bmatrix} x_1 & y_1 & 1 \\ x_T & y_T & 1 \\ x_2 & y_2 & 1 \end{bmatrix}$$

or

$$A = \underset{D}{\begin{bmatrix} 1 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix}} \begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_T & 0 \\ 0 & 0 & w_2 \end{bmatrix} \begin{bmatrix} x_1 & y_1 & 1 \\ x_T & y_T & 1 \\ x_2 & y_2 & 1 \end{bmatrix} \tag{4}$$

where matrix $D$ is the inverse of $B$.

The homogeneous coordinates $w_1$, $w_T$, and $w_2$ are unknown. We utilize the fourth point transformation $P_3 \rightarrow C_3$ to determine these quantities for a particular transformation. Let us illustrate this unique determination of $w_1$, $w_T$, $w_2$ by transforming the base parabola into a unit circle, shown in Figure 2, such that there is the correspondence

$$\begin{bmatrix} x_1 & y_1 & 1 \\ x_T & y_T & 1 \\ x_2 & y_2 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad C_3 = w_3[0 \quad 1 \quad 1]$$

(Note the correspondence $[x_T \quad y_T \quad 1] \Rightarrow [0 \quad 1 \quad 0]$, the point of intersection of tangents at $C_1$ and $C_2$, which is at infinity in the $y$ direction.)

Now

$C_3 = P_3A$ or

$$w_3[0 \quad 1 \quad 1]$$

$$= [\tfrac{1}{4} \quad \tfrac{1}{2} \quad 1] \begin{bmatrix} 1 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_T & 0 \\ 0 & 0 & w_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

which on solution yields

$$w_T = 2w_3 = w_1 = w_2$$

We arbitrarily set $w_3 = \frac{1}{2}$. (Reader may verify that this is harmless in homogeneous coordinates because $w_3$ is a common factor and could have any value.) Substitution of these values in Equation 4 yields

$$A = \begin{bmatrix} 0 & -2 & 2 \\ -2 & 2 & -2 \\ 1 & 0 & 1 \end{bmatrix}$$

which transforms the base parabola into the unit circle (center at origin).

By similar methods we can evaluate $A$ for any conic section. Hence, any conic section could be generated by the equation

$$[wx \quad wy \quad w] = [u^2 \quad u \quad 1] \begin{bmatrix} 3 \times 3 \\ \text{matrix} \end{bmatrix} \tag{5}$$

Specifically, we have evaluated the matrix that transforms four points from the plane of the parabola into the plane of the circle.

According to the fundamental theorem of plane perspectivity, four points in one coordinate system and four corresponding points in a transformed coordinate system completely define a projective transformation. Let us see if our matrix $A$, just obtained from the correspondence of four points, does indeed completely define the transformation of curves.

According to Equation 5, a general point on the circle should be

$$[wx \quad wy \quad w] = [u^2 \quad u \quad 1] \begin{bmatrix} 0 & -2 & 2 \\ -2 & 2 & -2 \\ 1 & 0 & 1 \end{bmatrix}$$

or the ordinary coordinates $x$, $y$ are given by

$$x = \frac{-2u + 1}{2u^2 - 2u + 1} \qquad y = \frac{-2u^2 + 2u}{2u^2 - 2u + 1}$$

(Reader may verify that $x^2 + y^2 = 1$.) Thus our Equation 5 is in conformance with the fundamental theorem of perspectivity.

**affine transformation**

A general affine transformation can be defined as one in which a space coordinate frame $0\ x\ y\ z$ is transformed into some other frame $0'\ x'\ y'\ z'$, generally speaking, with a different "metric," i.e., with unit segments of different lengths and with different angles between them, and in which a point $M$ is sent into point $M'$ having the same coordinates relative to the new frame as those of the point $M$ relative to the old frame[2] as illustrated in Figure 3. It can be shown that under an affine transformation every straight line is sent into a straight line, parallel lines are mapped into parallel lines, and if a point divides a segment in a given ratio, its image divides the image of this segment in the same ratio.

It follows from this definition that all circles and ellipses are affinely related to one another, i.e., one can be obtained from another by an affine transformation. In the last section, we obtained the parametric equation for a unit circle with its center at the origin. Now let us write the affine transformation to generate any circle of radius $r$ and center $(h, k)$ from the unit circle.

$$[wx \quad wy \quad w] = [u^2 \quad u \quad 1] \begin{bmatrix} 0 & -2 & 2 \\ -2 & 2 & -2 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} r & 0 & 0 \\ 0 & r & 0 \\ h & k & 1 \end{bmatrix}$$

The last matrix describes the scale change to radius $r$ followed by translation of the center to $(h, k)$. The reader may verify the result by performing the multiplication so that $(x - h)^2 - (y - k)^2 = r^2$, $(x, y)$ being the ordinary coordinates of a point on the circle. Furthermore, we can write the equation of any ellipse whose major axis is $2a$, whose minor axis is $2b$, whose center is at $(h, k)$, and whose major axis makes an angle $\Theta$ with the $x$ axis:

$$[wx \quad wy \quad w]$$

$$= [u^2 \quad u \quad 1] \underbrace{\begin{bmatrix} 0 & -2 & 2 \\ -2 & 2 & -2 \\ 1 & 0 & 1 \end{bmatrix}}_{\text{unit circle}} \underbrace{\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{scale}} \underbrace{\begin{bmatrix} \cos\Theta & \sin\Theta & 0 \\ -\sin\Theta & \cos\Theta & 0 \\ h & k & 1 \end{bmatrix}}_{\text{rotate-translate}}$$

We can show by similar methods that all hyperbolas are affine to one another, and so are all parabolas.

In passing, we remark that using the aforementioned methods for the generation of vectors for curves (conics in this case) eliminates the use of trigonometric functions and, hence, improves the "response time" for display.

Intersections of conic sections using these methods are discussed elsewhere.[3]

Figure 3 Affine transformatiom

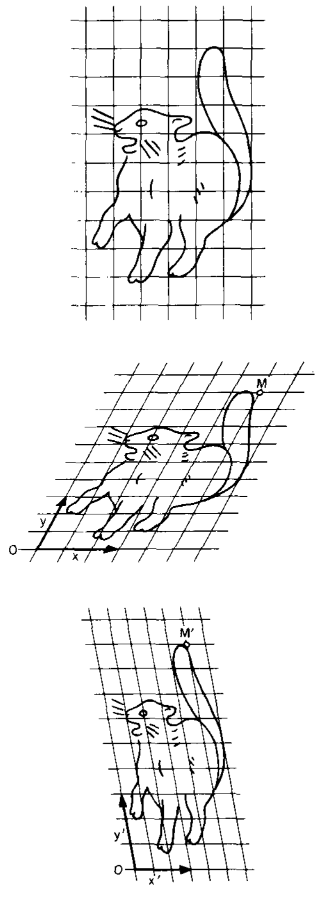


Figure 4 Coordinate system

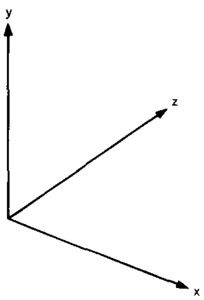



Figure 5 Top view by rotation

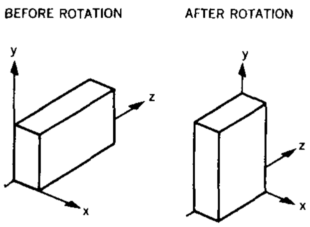
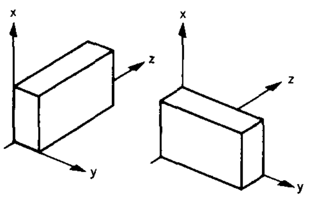



Figure 6 Right-side view by rotation




## Projection of three-dimensional objects into two-dimensional pictures

An important aspect of computer graphics is the projection of objects onto image planes.[4] All drawings and pictures are examples of special cases of projection. The orthographic, isometric, cavalier, and cabinet projections employed by draftsmen fall into this category, as do the perspective pictures used by architects. We can show that *all* of these two-dimensional images can be produced by a single $4 \times 4$ matrix, whose 16 elements are easy to determine.

Before we examine the general case, let us look briefly at the orthographic projections of an object as used by draftsmen—the ordinary top, front, and side views of an object. We establish a coordinate system such as in Figure 4 where $x$ is horizontal and increases to the right, $y$ is vertical and increases upward, and $z$ is horizontal and increases as indicated.

We shall think of the plane of $xy$, or the $z = 0$ plane, as the picture plane. Now imagine some object related to this coordinate system, and imagine that points on the object are represented by matrices of the form $[x \quad y \quad z]$. We might consider in particular a rectangular box with edges parallel to the three coordinate axes and with one corner at the origin of coordinates. There will be one corner of this box that lies in none of the three coordinate planes; let it be at $[1 \quad 2 \quad 3]$. Evidently the projection of this point on the picture plane is given by the point whose coordinates are $[1 \quad 2 \quad 0]$; that is, the $z$ coordinate after projection has become zero. The projection is given by the matric product

$$[1 \quad 2 \quad 3]\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = [1 \quad 2 \quad 0]$$

Now imagine that we wish to obtain a "top view" of this box. If $xy$ remains the picture plane, we must rotate the object with respect to this picture plane, and *after* the rotation we must project it into the picture plane.

**rotation**

Let us assume that rotation takes place as shown in Figure 5. The rotation is given by the matric product

$$[1 \quad 2 \quad 3]\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} = [1 \quad 3 \quad -2]$$

It consists of an interchange of the $y$ and $z$ coordinates, together with a sign change. We can also obtain a right-side view of the object by the rotation shown in Figure 6. Here the matric transformation is

$$[1 \quad 2 \quad 3]\begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = [3 \quad 2 \quad -1]$$

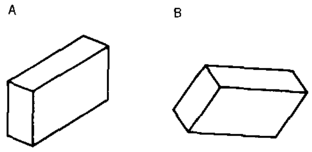
Again note the interchange of coordinates and the change of sign.

It works out that the determinants of both these rotation matrices are equal to $+1$. Without the change of sign, the determinants would be $-1$, which would be equivalent to producing a "reflection" of the object as well as a rotation. Right-hand objects would turn into left-hand objects. Indeed, it can be shown in general that the determinant of any rigid-rotation matrix has a value of $+1$, and this is a necessary (although not sufficient) condition on the matrix.

Finally, after either of the rotations, we obtain the projection on the $x$-$y$ plane by multiplying by the projection matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Figure 7 General projection of box




**axonometric pictorials**

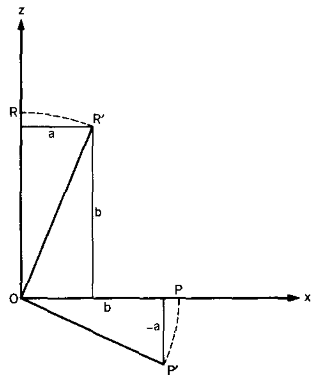
Now a *general* projection of an object can, as is well known, be obtained by drawing two auxiliary views of the object. This is essentially equivalent to making two arbitrary rotations in sequence, and then projecting the figure into the picture plane after the two rotations have been performed. The result will be an "axonometric pictorial" of the object.

Suppose we wish to produce a picture of an object in a general projection, but in addition, we wish to make vertical edges *appear* vertical after the transformation. We wish to achieve a picture of the box that looks like the one in Figure 7A and not like the one in Figure 7B.

We achieve the desired result by first rotating about the vertical $y$ axis. The rotation matrix is, in part,

$$\begin{bmatrix} & 0 & \\ 0 & 1 & 0 \\ & 0 & \end{bmatrix}$$

Figure 8 View from y axis




since $y$ dimensions do not change. But $x$ and $z$ locations change, and we need to examine the rotation to determine the appropriate matric entries.

Looking down the $y$ axis, we see the $x$ and $z$ axes as illustrated in Figure 8. Consider points $P$ and $R$ on each of these axes. If these points rotate rigidly about the origin $O$, they arrive at points $P'$ and $R'$. If the triangle $OPR$ is rigid, it is congruent to triangle $OP'R'$, and the right angle at $O$ is preserved, as well as the lengths. Let us say that the coordinates ($x$ and $z$) of point $R'$ after the rotation are $[a \quad b]$. If the coordinates of $R$ before rotation are $[0 \quad 1]$, and the length $OR$ = the length $OR'$, then $a^2 + b^2 = 1$; ($b$ can be the cosine of the angle of rotation, and $a$ can be the sine of this angle).

Again, the coordinates of $P$ before rotation can be $[1 \quad 0]$, and after rotation, the coordinates of $P'$ are necessarily $[b \quad -a]$ in order to preserve the right angle. When we examine this *plane* rotation, we have:

$$\begin{bmatrix} P \\ R \end{bmatrix} [T] = \begin{bmatrix} b & -a \\ a & b \end{bmatrix}$$

But

$$\begin{bmatrix} P \\ R \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and the rotation transformation matrix $T$ is then simply the matrix $\begin{bmatrix} b & -a \\ a & b \end{bmatrix}$ . Note that the determinant of this matrix is $+1$, since $a^2 + b^2 = 1$.

We introduce this result into the three-dimensional transformation to obtain

$$\begin{bmatrix} b & 0 & -a \\ 0 & 1 & 1 \\ a & 0 & b \end{bmatrix}$$

We now rotate the resulting figure about the $x$ axis; the $x$ coordinates do not change this time, and so the rotation matrix, in part, is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & & \\ 0 & & \end{bmatrix}$$

The missing partition is obtained as before and is equivalent to the plane rotation represented by the matrix $\begin{bmatrix} -d & c \\ c & d \end{bmatrix}$ where again $c^2 + d^2 = 1$, and $c$ and $d$ can be cosine and sine of the rotation angle. The complete three-dimensional rotation matrix is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -d & c \\ 0 & c & d \end{bmatrix}$$

The combination of rotations can be represented by the matric product of the separate primitive rotations taken in their proper order, and we evaluate it:

$$\begin{bmatrix} b & 0 & -a \\ 0 & 1 & 0 \\ a & 0 & b \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -d & c \\ 0 & c & d \end{bmatrix} = \begin{bmatrix} b & -ac & -ad \\ 0 & -d & c \\ a & bc & bd \end{bmatrix}$$

Observe the occurrence of the zero in the first (or new $x$-generating) column of the matrix. This zero occurs in the position that *will* be multiplied by the $y$ coordinate of the original point; it tells us that $x$ coordinates are independent of the heights of points on the object. A moment's reflection confirms that this statement is equivalent to saying that vertical lines appear vertical after the rotation, although they are certainly foreshortened. This is exactly what we want.

We now can achieve the projection, as before, by post-multiplication with the matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

**translation**

The foregoing procedure has left the corner of the object still attached to the origin. We wish to examine next the translation of the object to some new position in space. We now need homogeneous coordinates. If we wish to slide the object $e$ units to the right (in the $x$ direction), $f$ units upward (in the $y$ direction), and $g$ units back from the picture plane (in the $z$ direction), we can accomplish this by the transformation:

$$[x \quad y \quad z \quad 1]\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ e & f & g & 1 \end{bmatrix} = [(x+e) \quad (y+f) \quad (z+g) \quad 1]$$

The new point coordinates exhibit the translation. We can again compound the pure rotation transformation with this translation transformation, and we obtain the matric product

$$\left[\begin{array}{ccc|c} b & -ac & -ad & 0 \\ 0 & -d & c & 0 \\ a & bc & bd & 0 \\ \hline e & f & g & 1 \end{array}\right]$$

Notice that the rotation matrix and the translation matrix essentially appear in the compound matrix. The projection matrix is now

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Incidentally, the projection matrix is trivial; it represents a mathematical way of saying that we simply ignore the $z$ coordinates of the rotated and translated object when we construct the picture. But we see next that this matrix is nontrivial when we consider perspective pictorials, of which the preceding axonometric pictorials are a subclass.

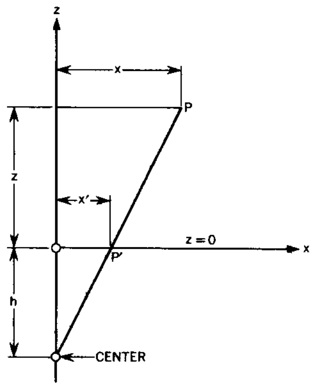
**perspective pictorials**

We next consider the projection of an object on the plane $z = 0$, but from a local point, say the point $[0 \quad 0 \quad -h]$ in ordinary coordinates. The situation can be pictured as in Figure 9. In this figure, the $y$ axis rises vertically out of the page.

We can write, by similar triangles, that

$$\frac{x'}{h} = \frac{x}{z + h},$$

Figure 9 Projection on plane z = 0




where $x'$ is the picture plane coordinate of $P'$, the image of $P$. This equation leads to

$$x' = \frac{xh}{z + h} = \frac{x}{(z/h) + 1}$$

A similar expression can be written for the

$$y' = \frac{y}{(z/h) + 1}.$$

Now consider the matric product:

$$[x \quad y \quad z \quad 1]\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/h \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \quad y \quad 0 \quad ((z/h) + 1)]$$

Obviously this matrix can be interpreted as equal to the homogeneous coordinate matrix

$$[wx \quad wy \quad 0 \quad w]$$

where $w = (z/h) + 1$, and

$$[x \quad y \quad 0 \quad 1] = (1/w)[wx \quad wy \quad 0 \quad w]$$

This relationship shows that the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/h \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

serves to project the object by rays from the center of projection, and the sectioning of this bundle of rays by the plane $z = 0$ produces the picture, which is accomplished by dividing the matrix $[wx \quad wy \quad 0 \quad w]$ by the quantity $w$.

We now see the nontrivial nature of the projection matrix and the need for homogeneous coordinates. Incidentally, if the distance $h$ from the center of projection to the picture plane is increased in the limit as $h$ approaches infinity, then $1/h$ approaches zero, and we obtain axonometric projection as a special case since the matrix becomes the trivial one already described.

The point $[0 \quad 0 \quad -h]$, the center of projection, is also the point from which the picture should be viewed. Any other viewing position yields more or less "distortion of perspective." This fact is very imperfectly understood, particularly by nontechnical artists and architects; this is obvious in much that has been written about distortion of perspectives and what empirical measures to take to avoid it. However, *any* perspective picture looks distorted unless viewed from this single point in space—but conversely any picture appears undistorted if this point is known and the picture is viewed from there.

distortion

If, in the construction of the picture, this point is, say, three inches from the picture, but the picture is viewed from a normal distance of, say, 15 inches, the perspective picture of necessity appears distorted, and violently so. It is difficult for most people to accommodate (or focus) the eye on a picture held three inches from the eye; however, if a person looks at the picture through a three-inch focal length magnifying glass, the picture will appear undistorted.

rational functions

We have seen that all conics can be generated by a transformation of a simple base conic by the formula, in homogeneous coordinates:

$$[wx \quad wy \quad w] = [u^2 \quad u \quad 1]\,A\,,$$

where $A$ is a $3 \times 3$ matrix. Then

$$x = \frac{wx}{w} \qquad \text{and} \qquad y = \frac{wy}{w}$$

Since $wx$, $wy$, and $w$ are each quadratic in $u$, we could call the coordinates $x$ and $y$ "rational quadratic functions" of $u$. Similarly, we can write[5]

$$[wx \quad wy \quad wz \quad w] = [u^3 \quad u^2 \quad u \quad 1]\,A$$

where $A$ is a $4 \times 4$ matrix.

In this case, the coordinates $x$, $y$, and $z$ are rational cubic functions of $u$. If the top row of $A$ consists of zeros, we have rational quadratics, or ordinary conics, a special case.

If the $A$ matrix is chosen so that its last column is

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

the curve is $[x \quad y \quad z \quad 1]$, and the denominator is always 1. This matrix represents an ordinary parametric cubic curve. Thus, conics and cubics are special cases of rational cubic functions, and a computer can generate circles, ellipses, hyperbolas, parabolas, and cubics, as well as more general curves, simply by proper choice of the $A$ matrix, without the necessity for having special and distinct routines for these curve forms. A specific application for generating spline-like curves utilizing this formula is discussed in this issue in a paper by Ahuja.

## Surfaces

A surface is the locus of a point that moves in space with two degrees of freedom. A point $V$ on a surface may be written in matric notation as:

$$[x \quad y \quad z] = [f(u, s) \quad g(u, s) \quad h(u, s)]$$

where $u$ and $s$ are independent parameters. Before proceeding further we shall compact the notation. We write

$$us \quad \text{for} \quad [f(u, s) \quad g(u, s) \quad h(u, s)]\,, \quad us_u \quad \text{for} \quad \frac{\partial(us)}{\partial u}\,,$$

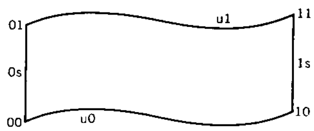
$$us_s \quad \text{for} \quad \frac{\partial(us)}{\partial s}\,, \quad us_{us} \quad \text{for} \quad \frac{\partial^2(us)}{\partial u \partial s}\,, \quad us_{uu} \quad \text{for} \quad \frac{\partial^2(us)}{\partial u^2}\,,$$

and likewise for other derivatives.

patches

We build complicated surfaces by adjoining small surface "patches." Accordingly, we focus our attention on one such surface patch. For computational simplicity, we restrict the variation of parameters in the range 0 to 1 for each patch, i.e., $0 \leq u, s \leq 1$. With this notation in mind, a surface patch can be considered to be a surface segment bounded by four space curves, $0s$, $1s$, $u0$, $u1$ as shown in Figure 10. (Note that symbol $u0$ stands for the vector describing the $(x \quad y \quad z)$ coordinates of points along the curve generated by holding $s = 0$ constant and varying $u$.) We wish to blend such patches (for example, $A1$ and $A2$ in Figure 11) into one surface with any desired characteristics at common boundaries. The surface equation for a slope-matching, slope-continuous surface patch with entirely arbitrary boundaries and entirely arbitrary slopes across these boundaries may be written in matric notation:[5]

**Figure 10** Surface patch

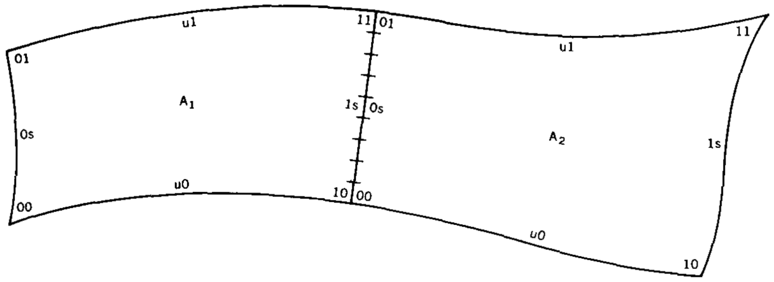



$$us = -[-1 \quad F_0u \quad F_1u \quad G_0u \quad G_1u]$$
$$\times \left[\begin{array}{c|cc|cc} 0 & u0 & u1 & u0_s & u1_s \\ \hline 0s & 00 & 01 & 00_s & 01_s \\ 1s & 10 & 11 & 10_s & 11_s \\ \hline 0s_u & 00_u & 01_u & 00_{us} & 01_{us} \\ 1s_u & 10_u & 11_u & 10_{us} & 11_{us} \end{array}\right] \left[\begin{array}{c} -1 \\ F_0s \\ F_1s \\ G_0s \\ G_1s \end{array}\right] \tag{6}$$

where $F_0$, $F_1$, $G_0$, $G_1$, are scalar functions of a single variable with the following end conditions:

$$F_0(0) = F_1(1) = 1\,,$$
$$F_0(1) = F_1(0) = G_0(0) = G_1(0) = G_1(1) = G_0(1) = 0\,,$$
$$F_0'(0) = F_1'(0) = F_1'(1) = F_0'(1) = G_0'(1) = G_1'(0) = 0\,,$$
$$G_0'(0) = G_1'(1) = 1$$

**Figure 11 Surface patches to be blended**




These functions serve to blend the aforementioned characteristics in the surface patch and are hence called *blending* functions. Equation 6 can be easily expanded for higher derivatives continuity.[5] Blending functions with the previously described stipulations can be used to define curves in terms of their end points and end-point tangent vectors, e.g.,

$$u0 = [F_0u \quad F_1u \quad G_0u \quad G_1u] \begin{bmatrix} 00 \\ 10 \\ 00_u \\ 10_u \end{bmatrix}$$

Furthermore, we can relate the blending function vector to a so-called *basis vector* $[u_1 \quad u_2 \quad u_3 \quad u_4]$ in the following way:

$$[F_0u \quad F_1u \quad G_0u \quad G_1u] = [u_1 \quad u_2 \quad u_3 \quad u_4]\, M$$

With an appropriate choice of the basis vector, Equation 6 can be used to generate a very wide class of surfaces. Specifically, if the basis vector is chosen to be $[u^3 \quad u^2 \quad u \quad 1]$, if $u0$, $u1$, $u0_s$, $u1_s$, $0_s$, $1s$, $0s_u$, $1s_u$ are linear combinations of the elements of the basis vector, and if the expression is written in homogeneous form, then we obtain a $4 \times 4 \times 4$ tensor as descriptive of the boundary conditions. As a special case, this tensor leads to a parametric description of quadric surfaces.

## CITED REFERENCES

1. J. T. Rule and R. F. Watts, *Engineering Graphics,* McGraw-Hill Book Company, New York, New York, 49 (1951).
2. A. D. Alexendrov, A. N. Kolmgrove and M. A. Lavrent'ev, *Mathematics: Its Content, Methods and Meaning,* American Mathematical Society, New York, New York (1963).
3. L. G. Roberts, *Homogeneous Matrix Representation and Manipulation of N-Dimensional Constructs,* Document MS 1405, Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts (May 1965).
4. L. G. Roberts, *Machine Perception of Three-Dimensional Solids,* Technical Report No. 315, Massachusetts Institute of Technology, Cambridge, Massachusetts (May 1963).
5. S. A. Coons, *Surfaces for Computer-Aided Design of Space Forms,* MAC-TR-41, Clearinghouse for Federal, Scientific, and Technical Information, Springfield, Virginia (June 1967).