tanulmányok 73/1978

MTA Számitástechnikai és Automatizálási Kutató Intézet Budapest

MAGYAR TUDOMÁNYOS AKADÉMIA
SZÁMITÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZETE

HOMOGENEOUS COORDINATES, PROJECTIVE TRANSFORMATIONS, AND CONICS

TUTORIAL

by

S.A. Coons

Tanulmányok 73/1978.

A kiadásért felel:

DR VÁMOS TIBOR

ISBN 963 311 056 4

Készült:
az ORSZÁGOS MŰSZAKI KÖNYVTÁR ÉS DOKUMENTÁCIÓS KÖZPONT
házi sokszorosítójában
F. v.: Janoch Gyula

HOMOGENEOUS COORDINATES, PROJECTIVE TRANSFORMATIONS, AND CONICS

TUTORIAL

by

S. A. Coons

Computer and Automation Institute

Hungarian Academy of Sciences

The equation for a line in two dimensions,

$$Ax + By + C = 0$$

is, in matric form,

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

But we observe that

$$[\,hx \quad hy \quad h\,] \begin{bmatrix} A \\ B \\ C \end{bmatrix} = 0 \quad \text{also.}$$

Here h is any multiplier whatever, of the vector $[x \quad y \quad 1]$.

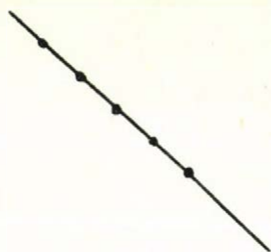
Evidently again

$$[\,hx \quad hy \quad h\,] \begin{bmatrix} \alpha A \\ \alpha B \\ \alpha C \end{bmatrix} = 0 \text{ also, for any } \alpha.$$

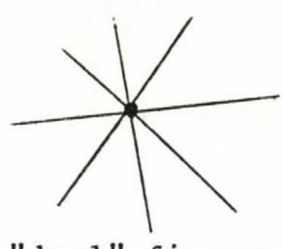
We call $[\,hx \quad hy \quad h\,]$ a point vector, and $\begin{bmatrix} \alpha A \\ \alpha B \\ \alpha C \end{bmatrix}$ a line vector.

The expression $[hx \quad hy \quad h] \begin{bmatrix} A \\ B \\ C \end{bmatrix} = 0$ is a statement that the point $[hx \quad hy \quad h]$ lies on the line $\begin{bmatrix} \alpha A \\ \alpha B \\ \alpha C \end{bmatrix}$.

If we regard $[hx \quad hy \quad h]$ as a variable vector, and $\begin{bmatrix} \alpha A \\ \alpha B \\ \alpha C \end{bmatrix}$ as a fixed set of constants, then we describe points that lie on a fixed line. Pictorially, this is:



Contrariwise, when we regard $[hx \quad hy \quad h]$ as a fixed vector, with three fixed constants, and $\begin{bmatrix} \alpha A \\ \alpha B \\ \alpha C \end{bmatrix}$ as a variable vector, then we describe lines that pass through a fixed point. Pictorially, this is:



These two figures are "dual" figures.

We say in words that the first figure is the "point row" that lies on a line, and the second figure is the "ray-sheaf", that lies on a point.

Sometimes we say that the line is the "support" for the point-row, and the point is the "support" for the ray-sheaf.

In either case, the products

$$[hx \quad hy \quad h] \begin{bmatrix} \alpha A \\ \alpha B \\ \alpha C \end{bmatrix} = 0$$

and $$[\alpha A \quad \alpha B \quad \alpha C] \begin{bmatrix} hx \\ hy \\ h \end{bmatrix} = 0$$

are algebraic expressions of these two dual geometric figures.

- . - . -

We choose the letter "h" in $[hx \quad hy \quad h]$ to indicate what we will call the "homogeneous coordinate" of the point $[hx \quad hy \quad h]$.

In the equation

$$[hx \quad hy \quad h] \begin{bmatrix} \alpha A \\ \alpha B \\ \alpha C \end{bmatrix} = 0,$$ it doesn't matter what the numerical value of h is.

For any value of h except $h = 0$, we have the coordinates x and y, that we can find from

$$x = \frac{hx}{h} \quad \text{and} \quad y = \frac{hy}{h}.$$

But what happens in case $h = 0$?

Then

$$x = \frac{0x}{0} \quad \text{and} \quad y = \frac{0y}{0}$$ are meaningless quantities.

This is just algebra trying to tell us that there are no finite numbers x and y that satisfy these two equalities.

Nevertheless, as will soon be seen, we can have vectors for points such as $[hx \quad hy \quad 0]$.

In such a case, we regard hx and hy as proper numbers, but we cannot find finite solutions for x and y.

In other words, if $[hx \quad hy \quad h] = [a \quad b \quad 0]$, then this must represent a special point, a "point of infinity", and this is an example of the usefulness of the matric notation, and of homogeneous coordinates.

(Far from being a disaster that there is no x that satisfies the equation $x = \frac{a}{0}$, it points the way to describing points at infinity in finite arithmetical ways.)

To obtain some intuitive notion of the concept of a "point at infinity", consider the point vector $[hx \quad hy \quad h]$.

When h = 1, this is simply $[x \quad y \quad 1]$.
Now, without changing hx and hy, consider $[hx \quad hy \quad \frac{h}{10}]$.

Then $x = \frac{hx}{h} \times 10$ and $y = \frac{hy}{h} \times 100$.

When the third component is $\frac{h}{100}$, still smaller, then

$$x = \frac{hx}{h} \times 100 \qquad y = \frac{hy}{h} \times 100.$$

And so forth.
Evidently, in the limit, the point is represented by $[hx \quad hy \quad 0]$, with x and y "infinitely" large. That is to say, the "ordinary" coordinates x and y are not finite quantities. We look only at the biliteral quantities hx and hy, which, strangely enough, are not zero. This is to say that a vector $[a \quad b \quad 0]$ represents a point at infinity.

It will be seen that $[a \quad b \quad 0]$ consists of a set of three numbers, not all of them zero, when we perform a transformation.

- . - . - . -

We form the matric product

$$[hx \quad hy \quad h] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = [h^*x^* \quad h^*y^* \quad h^*],$$

three new components of what we suspect is a new, transformed point vector. By analogy with

$$x = \frac{hx}{h} \quad \text{and} \quad y = \frac{hy}{h}, \text{ we suspect that}$$

$$x^* = \frac{h^*x^*}{h^*} \quad \text{and} \quad \frac{h^*y^*}{h^*},$$ (except of course when $h^* = 0$.)

We can abbreviate the notation, writing $hvT = h^*v^*$ where T is the square transformation matrix, a_{ij}.

We have, for the point-line coincidence,

$[hv]\ [\alpha L] = 0$ where hv is a point vector, and αL is a line vector.

$$hv = [hx \quad hy \quad h] \quad \text{and} \quad \alpha L = \begin{bmatrix} \alpha A \\ \alpha B \\ \alpha C \end{bmatrix}.$$

We can now set

$[hv]\ TT^{-1}\ [\alpha L] = 0$, since $TT^{-1} = I$, which is the "identity" matrix.

$$hv\ T = h^*v^*, \quad \text{and} \quad T^{-1}\ L = \alpha^*L^*, \text{ so}$$

$$[h^*v^*]\ [\alpha^*L^*] = 0.$$

We can interpret h^*v^* as a transformed point, and α^*L^* as a transformed line, and the transformed point lies on the transformed line. More compactly, if p = point vector, λ = line vector, and if $p\lambda = 0$, then

$p\ TT^{-1}\lambda = 0, \qquad (p\ T)\ (T^{-1}\lambda) = 0,$

$p\ T = p^*, \qquad T^{-1}\lambda = \lambda^*$, and there results

$p^*\lambda^* = 0.$ This is again a "linear form."

- . - . - . -

The foregoing indicates that points that lie on a line in the "object"space, after the transformation T, lie on a line in the "image" space.

Collinear points transform into collinear points.
(The dual: concurrent lines transform into concurrent lines.)

We "know", that "two points determine a line." Let us see what this means algebraically

$$[h_1x_1 \quad h_1y_1 \quad h_1] \begin{bmatrix} A \\ B \\ C \end{bmatrix} = 0$$ is a statement

that the point lies on the line.

For <u>two</u> points, we have

$$\begin{bmatrix} h_1x_1 & h_1y_1 & h_1 \\ h_2x_2 & h_2y_2 & h_2 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$ and

we wish to find an appropriate set of quantities A, B, C, that satisfy this matric equation.

For notational simplicity, we write

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Now, changing the subject for a moment, observe (as Cramer no doubt did) that always

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix} = 0.$$

This is a determinant whose value is zero, since two rows are identical.

Expanding, it is guaranteed that

$$a_1 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} - b_1 \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} + c_1 \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = 0$$

But also

$$\begin{vmatrix} a_2 & b_2 & c_2 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = a_2 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} - b_2 \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} + c_2 \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

It is seen that the quantities

$$\begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}, \quad - \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

are a plausible set of numbers A, B, and C such that

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

- . - . -

It is interesting in passing to observe that these numbers are just the components of the so-called <u>vector product</u> of $[x \;\; y \;\; 1]$ and $[x_2 \;\; y_2 \;\; 1]$.

There is more significance to this remark than might be seen at first glance.

The scalar product $[a \;\; b \;\; c] \begin{bmatrix} A \\ B \\ C \end{bmatrix} = 0$.

<u>also</u> means that in the three-dimensional space, the vector

[a b c] is orthogonal to the vector [A B C]. Thus the matric equation

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

describes a vector [A B C] that is mutually orthogonal to both $[a_1\ b_1\ c_1]$ and $[a_2\ b_2\ c_2]$. As a consequence of the single notion of "scalar product" of two vectors, we arrive at the notion of "vector product," which is now seen to be derivative.

["Scalar product" is sometimes called the "dot product" or "inner product" of two vectors.]

We can also have

$$[hx \quad hy \quad h] \begin{bmatrix} A_1 & A_2 \\ B_1 & B_2 \\ C_1 & C_2 \end{bmatrix} = [0 \quad 0]$$

This is an equation that describes the point of intersection of two lines.

[Thus: "Two points determine a line" and
"Two lines determine a point".]

These are dual remarks, and the algebra is clear.

– . – . –

As an interesting special case, consider <u>all</u> points at infinity, [a b 0], where a and b are <u>any</u> numbers, <u>as long as not both are zero</u>.

Now it is guaranteed that

$$[a \quad b \quad 0] \begin{bmatrix} 0 \\ 0 \\ C \end{bmatrix} = 0. \text{ For any a, b, and C.}$$

It is just as good to write

$$\begin{bmatrix} 0 \\ 0 \\ C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{if we wish.}$$

This is just a way of saying that all points at infinity in the plane lie on the single line whose vector is

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

It is, perhaps, hard to think that there is only one line at infinity in the plane, but this is what the arithmetic insists is true.

– . – . –

We need now to turn our attention to the details of transformations. We will consider three fragments of the general transformation: first,

$$[hx \quad hy \quad h] \left[\begin{array}{cc|c} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ \hline 0 & 0 & 1 \end{array}\right] = [h^*x^* \quad h^*y^* \quad h^*]$$

Next, we will consider

$$[hx \quad hy \quad h] \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline a_{31} & a_{32} & 1 \end{array}\right] = [h^*x^* \quad h^*y^* \quad h^*]$$

and finally, we consider

$$[hx \quad hy \quad h] \left[\begin{array}{cc|c} 1 & 0 & a_{13} \\ 0 & 1 & a_{23} \\ \hline 0 & 0 & a_{33} \end{array}\right] = [h^*x^* \; h^*y^* \; h^*] .$$

The first transformation gives

$$[hx \quad hy \quad h] \left[\begin{array}{cc|c} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ \hline 0 & 0 & 1 \end{array}\right]$$

$$= [hx^* \quad hy^* \quad h] , \qquad \text{since } h^* = h.$$

Since the third column and the third row of this transformation do nothing, we can write

$$[x \quad y] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = [x^* \quad y^*].$$

This is the so-called "affine transformation" in the plane. Now consider some special values of $[x \quad y]$:

$$\left[\begin{array}{cc} 0 & 0 \\ \hline 1 & 0 \\ 0 & 1 \end{array}\right] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \left[\begin{array}{cc} 0 & 0 \\ \hline a_{11} & a_{12} \\ a_{21} & a_{22} \end{array}\right]$$

This equation tells us that the origin of coordinates, $[0 \quad 0]$ transforms into <u>itself.</u> But the unit point on the x-axis, $[1 \quad 0]$, transforms into the point $[a_{11} \quad a_{12}]$, and the unit point on the y axis transforms into the point $[a_{21} \quad a_{22}]$.

This is simply a way of saying that the transformation matrix can be written directly from

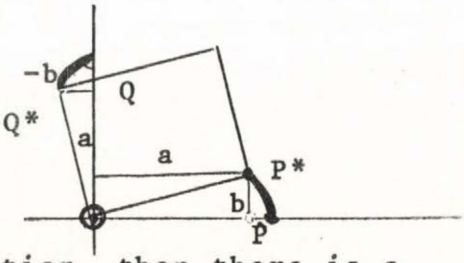
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

where $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is the matrix describing the unit points on the y axis (the "object" points,) and

$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is the matrix describing their images.

- . - . -

A "pure" rotation is described by a special case of such a matrix. The coordinates for the unit point P on the x axis are $[1 \quad 0]$ and after the rotation, the coordinates are $[a \quad b]$ for P*.



If the transformation is a "pure" rotation, then there is a length-preserving condition, which says that

$$|O\,P| = |O\,P^*| .$$

$$|O\,P| = 1 , \quad \text{and} \quad |O\,P^*| = \left([a \quad b] \begin{bmatrix} a \\ b \end{bmatrix} \right)^{1/2} = 1$$

So $a^2 + b^2 = 1$.

The unit point Q transforms into Q* whose coordinates are $[-b \quad a]$.

The complete rotation transformation is

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

For $a^2 + b^2 = 1$, we can take $a = \cos\,\Theta$, $b = \sin\,\Theta$ and

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \cos\,\Theta & \sin\,\Theta \\ -\sin\,\Theta & \cos\,\Theta \end{bmatrix}$$, the familiar

transformation matrix, well known from analytic geometry.

The matrix $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ is called an "ortho-normal" matrix. The product of it and its transpose is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, the identity matrix.

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \left[\begin{array}{c|c} a^2 + b^2 & ab - ab \\ ab - ab & a^2 + b^2 \end{array}\right], \quad \text{and since}$$

$$a^2 + b^2 = 1, \quad \text{this is} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

But this is just a way of saying that the transpose of an ortho-normal matrix is also the inverse of the matrix (and of course conversely.)

We have just investigated the partitioned general matrix, and found the significance of the fragment

$$\left[\begin{array}{cc|c} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ \hline 0 & 0 & 1 \end{array}\right].$$

Now consider another matrix transformation,

$$[hx \quad hy \quad h] \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline a_{31} & a_{32} & 1 \end{array}\right] = [hx + ha_{31} \quad hy + ha_{32} \quad h]$$

which, if $h \neq 0$, becomes, after dividing by h,

$$[x + a_{31} \quad y + a_{32} \quad 1] .$$

This means that a point $[x \quad y \quad 1]$ undergoes a simple displacement in the plane, by the displacement vector $[a_{31} \quad a_{32} \quad 1]$.

The result of an affine (possibly a pure rotation) transformation followed by a displacement transformation is the transformation product:

$$\left[\begin{array}{cc|c} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ \hline 0 & 0 & 1 \end{array}\right] \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline a_{31} & a_{32} & 1 \end{array}\right] = \left[\begin{array}{cc|c} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ \hline a_{31} & a_{32} & 1 \end{array}\right]$$

Observe in passing that the origin of planar coordinates, $[0 \quad 0 \quad 1]$, becomes the new transformed point $[a_{31} \quad a_{32} \quad 1]$. This makes it possible to "slide things around" in the plane.

- . - . -

Finally consider the matrix

$$\left[\begin{array}{cc|c} 1 & 0 & a_{13} \\ 0 & 1 & a_{23} \\ \hline 0 & 0 & a_{33} \end{array}\right]$$

where the affine partition is just the identity partition, and the displacement partition likewise does no displacement, but where the right-hand partition does "something" to the image. We will call this matrix the "projection" matrix, and assign the symbol P to it.

To learn the significance of this matrix, consider the equation of point-line:

$$v\ L = 0$$

Then $v\ PP^*L = 0$, where P^* is the "adjoint" of P, such that

$$PP^* = \begin{bmatrix} \delta & 0 & 0 \\ 0 & \delta & 0 \\ 0 & 0 & \delta \end{bmatrix}$$

If $\delta \neq 0$, then obviously

$$v\ PP^*L = v\ \delta \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} L = 0.$$

But even if $\delta = 0$, the equality still holds, obviously.

The adjoint of P is P*,

$$P^* = \left[\begin{array}{cc|c} a_{33} & 0 & -a_{13} \\ 0 & a_{33} & -a_{23} \\ \hline 0 & 0 & 1 \end{array}\right], \qquad \text{which we verify:}$$

$$PP^* = \left[\begin{array}{cc|c} 1 & 0 & a_{13} \\ 0 & 1 & a_{23} \\ \hline 0 & 0 & a_{33} \end{array}\right] \left[\begin{array}{cc|c} a_{33} & 0 & -a_{13} \\ 0 & a_{33} & -a_{23} \\ \hline 0 & 0 & 1 \end{array}\right]$$

$$= \begin{bmatrix} a_{33} & 0 & 0 \\ 0 & a_{33} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} = a_{33} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Now the new, transformed, line vector is

$$P^* L = L^* .$$

Thus, presumably, for every line L in the object space, there is a corresponding line L^* in the image space.

But what if $a_{33} = 0$? Then for $L = \begin{bmatrix} A \\ B \\ C \end{bmatrix}$

$$P^* L = \left[\begin{array}{cc|c} 0 & 0 & -a_{13} \\ 0 & 0 & -a_{23} \\ \hline 0 & 0 & 1 \end{array}\right] \begin{bmatrix} A \\ B \\ C \end{bmatrix} \qquad \text{and the transformed line}$$

$$L^* = \begin{bmatrix} -a_{13} & C \\ -a_{23} & C \\ & C \end{bmatrix}.$$

Since we are dealing with homogeneous coordinates, this can be written as $\begin{bmatrix} -a_{13} \\ -a_{23} \\ 1 \end{bmatrix}$, a unique line vector that represents all possible transformations of lines $\begin{bmatrix} A \\ B \\ C \end{bmatrix}$ in the object space. All lines in the object space transform into a single line in the image space.

This is interesting, because the transformation matrix

$$P^* = \left[\begin{array}{cc|c} 0 & 0 & -a_{13} \\ 0 & 0 & -a_{23} \\ 0 & 0 & 1 \end{array}\right]$$

is evidently a singular matrix.

This says that when once the transformation has happened, there is no way to recover the object line. All lines in object space are indistinguishable in this image space. There's just no going back.

– . – . –

As we shall soon see, there exists an analogous transformation that similarly transforms all planes of the object space into a single plane in the image space, when we begin to discuss three-dimensional transformations. Sometimes people naively assume that such a transformation describes the behavior of a photographic camera, in which three-dimensional objects are imaged on a two-dimensional film. Such is not the case. We will have a little more to say on the subject later, when we examine the photographic transformation in the light of the geometry of 3-dimensional transformations.

For the time being, however, we must return to two-dimensional transformations, because there are some quite interesting results that we can achieve.

We have looked at the geometric interpretations of three transformation fragments:

1. The affine transformation that leaves the origin of the coordinate system fixed.

2. The displacement transformation that "slides" everything elsewhere - a "pure translation" without rotation, expansion, or other changes.

3. The pure projective transformation, that transforms object points into image points in a very interesting way.

Call these transformations the A, D, P transformations. If we form the product of the matric fragments, in the order $A\ D\ P = T$, then the compound matrix T will consist of nine numbers; it will be a full 3×3 matrix.

We are now ready to show that <u>four points</u> in the object space and their four images in the image space completely define a projective transformation.

For the points in the object space, we may use the generic form $[x\ \ y\ \ 1]$ if they are local points, or $[a\ \ b\ \ 0]$ if they are points at infinity. In either case, represent these object points with the symbol $\underline{r}$ so that three points are represented by the matrix $\begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}$. (While this looks like a simple column vector, it is in reality a 3x3 square matrix.)

For the points in the image space, we use the generic biliteral symbol hv. Here typically

$$hv = [hx \quad hy \quad h] \quad \text{where } x \text{ and } y$$

are "ordinary" coordinates, and h is the third homogeneous coordinate.

We can now write the transformation equation, using these symbolic abbreviations:

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} T = \begin{bmatrix} h_1v_1 \\ h_2v_2 \\ h_3v_3 \end{bmatrix} .$$

The square 3x3 matrix on the right can be factored, and we have

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} T = \begin{bmatrix} h_1 & & \\ & h_2 & \\ & & h_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

(We omit the zero's in the h_i diagonal matrix so as to make it appeal more directly to the eye.)

We can now partly solve for the transformation matrix T:

$$T = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}^{-1} \begin{bmatrix} h_1 & & \\ & h_2 & \\ & & h_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} .$$

The r_i and the v_i are the known coordinate vectors of object points and their corresponding images. But the three quantities h_1 h_2 and h_3 are still unknown.

We need one additional piece of information, and we introduce still another point in the object space, r_4, and its image, v_4.

Then

$$r_4 \, T = h_4 \, v_4 .$$

In greater detail, this is

$$r_4 \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}^{-1} \begin{bmatrix} h_1 & & \\ & h_2 & \\ & & h_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = h_4 v_4 .$$

We now take the liberty of setting $h_4 = 1$. This is the same as saying that we allow h_4 to be "absorbed" by the three unknown place-holders h_1 h_2 and h_3.

We again take an inverse, and obtain

$$r_4 \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}^{-1} \begin{bmatrix} h_1 & & \\ & h_2 & \\ & & h_3 \end{bmatrix} = v_4 \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}^{-1}$$

Notice that

$$r_4 \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}^{-1} = [a \quad b \quad c]$$, a vector, a 1x3 matrix.

Likewise,

$$v_4 \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}^{-1} = [A \quad B \quad C]$$, another vector.

We now have

$$[a \quad b \quad c] \begin{bmatrix} h_1 & & \\ & h_2 & \\ & & h_3 \end{bmatrix} = [A \quad B \quad C]$$

where everything is known except for the h_i diagnoal matrix. But now we can write

$$[a\,h_1 \quad b\,h_2 \quad c\,h_3] = [A \; B \; C] \quad ,$$

from which

$$h_1 = \frac{A}{a}$$

$$h_2 = \frac{B}{b}$$

$$h_3 = \frac{C}{c}$$

Thus the fourth point r_4 and its image v_4 enable us to solve for the h_i quantities, which we can introduce into the equation for the T transformation matrix, to define T completely.

- . - . -

AN EXAMPLE

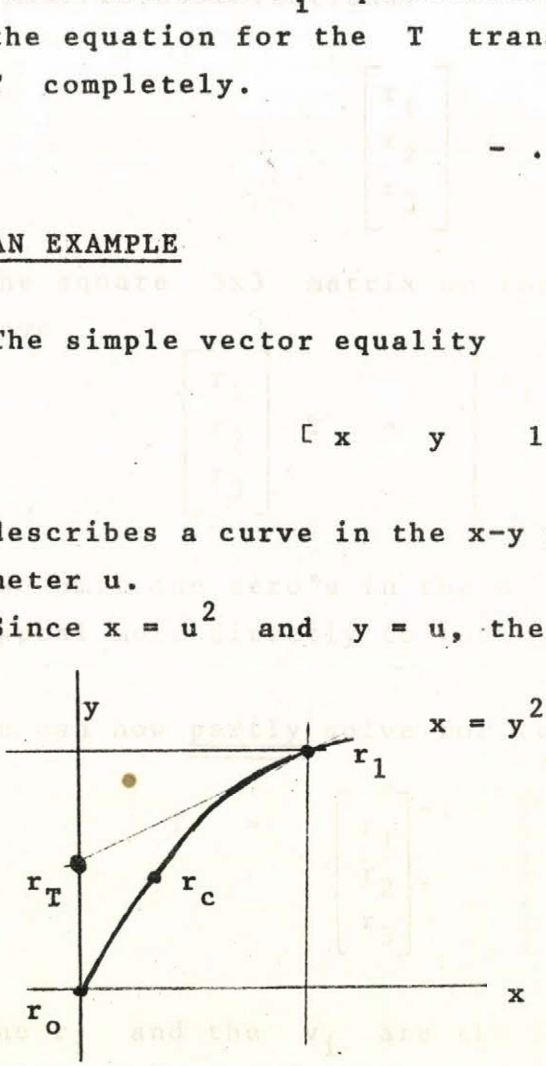
The simple vector equality

$$[\,x \quad y \quad 1\,] = [\,u^2 \quad u \quad 1\,]$$

describes a curve in the x-y plane, as a function of the parameter u.

Since $x = u^2$ and $y = u$, the equation for this curve is

$x = y^2$ a simple parabola.



We take four points in this plane,

$$\begin{bmatrix} r_o \\ r_1 \\ r_c \end{bmatrix} \quad \text{points corresponding to} \quad u = 0 \quad u = 1 \quad u = 1/2$$

and the point r_T where the two tangents to the parabola intersect. These two tangents touch the curve at r_o and at r_1.

It is easy to write the values of the r vectors. They are

$$\begin{bmatrix} r_o \\ r_T \\ r_1 \\ \hline r_c \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1/2 & 1 \\ 1 & 1 & 1 \\ \hline 1/4 & 1/2 & 1 \end{bmatrix}$$

For r_c we can use $[1 \quad 2 \quad 4]$ instead of $[1/4 \quad 1/2 \quad 1]$.

Let these four points be transformed into four different points, $\begin{bmatrix} v_o \\ v_T \\ v_1 \\ v_c \end{bmatrix}$ which we can choose arbitrarily.

We have already learned that

$$r_c \begin{bmatrix} r_o \\ r_T \\ r_1 \end{bmatrix}^{-1} \begin{bmatrix} h_o & & \\ & h_T & \\ & & h_1 \end{bmatrix} = v_c \begin{bmatrix} v_o \\ v_T \\ v_1 \end{bmatrix}^{-1}$$

and this equation serves to enable us to find the h_i quantities.

Now

$$r_c \begin{bmatrix} r_o \\ r_T \\ r_1 \end{bmatrix}^{-1} = [1 \quad 2 \quad 4] \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1/2 & 1 \\ 1 & 1 & 1 \end{bmatrix}^{-1}$$

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

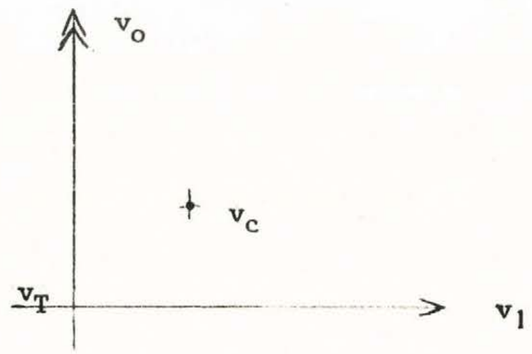
This is a symmetric matrix, and when we come to look at Bézier curves, we will see that it is one of the Bézier matrices.

Multiplying out, we find that

$$r_c \begin{bmatrix} r_o \\ r_T \\ r_1 \end{bmatrix}^{-1} = [1 \quad 2 \quad 4] \begin{bmatrix} 1 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix} = [1 \quad 2 \quad 1]$$

For the v_i points of the image, let us take

$$\begin{bmatrix} v_o \\ v_T \\ v_1 \\ \hline v_c \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ \hline 1 & 1 & 1 \end{bmatrix}$$



The sketch is intended to suggest that point v_o is at infinity on the y axis, v_1 is at infinity on the x axis, v_T is at the origin of coordinates, and v_c is at $x = 1$, $y = 1$, all of these points being expressed in homogeneous coordinates, of course.

Now
$$\begin{bmatrix} v_o \\ v_T \\ v_1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

and
$$v_c \begin{bmatrix} v_o \\ v_T \\ v_1 \end{bmatrix}^{-1} = [1 \quad 1 \quad 1]$$

We now have the general expression

$$[a\, h_o \quad b\, h_T \quad c\, h_1] = [\, A \quad B \quad C \,].$$

which gives immediately

$$[\, h_o \quad 2\, h_T \quad h_1 \,] = [\, 1 \quad 1 \quad 1 \,]$$

or $$[\, h_o \quad h_T \quad h_1 \,] = [1 \quad 1/2 \quad 1] \text{ or } [2 \quad 1 \quad 2],$$

to avoid the fraction.

We now have all the ingredients of the transformation matrix T:

$$T = \begin{bmatrix} r_o \\ r_T \\ r_1 \end{bmatrix}^{-1} \begin{bmatrix} h_o & & \\ & h_T & \\ & & h_1 \end{bmatrix} \begin{bmatrix} v_o \\ v_T \\ v_1 \end{bmatrix}$$

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

$$= \begin{bmatrix} 2 & 2 & -2 \\ 0 & -4 & 2 \\ 0 & 2 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

when we divide out the common factor 2.

First, let us test to see whether this matrix does indeed transform the r_i points into the v_i points.

<u>$r_i\, T = v_i$</u>

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

$$[1 \quad 2 \quad 4] \qquad = \quad [1 \quad 1 \quad 1] \qquad [1 \quad 1 \quad 1]$$

for all four points.

We now have a formula for some kind of curve, whose vector equation is

$$[hx \quad hy \quad h] = [u^2 \quad u \quad 1]\begin{bmatrix} 1 & 1 & -1 \\ 0 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$ It is

a transformation of the primitive parabola.

What kind of curve is it?

Consider the identity

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

$$= u^2 - u^2 \equiv 0$$

The singular matrix $\begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ which we will call

the "k matrix" has put in a gratuitous appearance, which we shall explain later.

For the primitive parabolic function,

$$[x \quad y \quad 1] = [u^2 \quad u \quad 1],$$ we can write

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

$$= x - y^2 = 0.$$

Abbreviate this to

$$r\,k\,r^T \doteq 0\,.$$

The transformed curve is

$$r\,T = h\,v\,.$$

We can now write harmlessly,

$$r \quad T \, T^{-1} \, k \, T^{-1T} \, T^{T} \, r^{T} \quad = \quad 0$$

because

TT^{-1} and $T^{-1T} \, T^{T}$ both yield identity matrices.

But now

$$(r \, T) \quad (T^{-1} \, k \, T^{-1T}) \quad (T^{T} \, r^{T}) \quad = 0$$

or

$$hv \quad (T^{-1} \, k \, T^{-1T}) \quad hv^{T} \quad = 0$$

Call the product $(T^{-1} \, k \, T^{-1T}) = C$, a matrix.

Then

$$hv \, C \, hv^{T} = 0 \qquad \text{or} \qquad v \, C \, v^{T} = 0,$$

since $h = 0$ yields a triviality.

This is known as a quadratic form.

For our immediate example, let us evaluate C.

First,

$$T^{-1} = \begin{bmatrix} 1 & 1 & -1 \\ 0 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \overline{T}$$

Then

$$\overline{T} \, k \, \overline{T}^{T} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

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

This gives the quadratic form,

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

We multiply as indicated:

$$[x \quad y \quad 1]\begin{bmatrix} y+1 \\ 0 \\ -x-1 \end{bmatrix} = xy + x - x - 1 =$$

$$= xy - 1 = 0 .$$

This comes out to be $xy = 1$ or $y = \frac{1}{x}$, the simple hyperbola.

So we have been able to transform the simple primitive parabola $x = y^2$ into the simple hyperbola $y = \frac{1}{x}$.

Before leaving this, consider the value of the independent parameter, the "driving" variable, when $u = \infty$, (increases without limit.) For the primitive conic, the parabola $[u^2 \quad u \quad 1]$ becomes $[1 \quad 0 \quad 0]$ since when u is immensely large, and growing, then u^2 is immensely large with respect to u, and certainly with respect to 1. But we are dealing with the benign properties of homogeneous vector quantities, which allow us to treat of infinities.

For the primitive parabola, $[x \quad y \quad 1] = [u^2 \quad u \quad 1]$, when u increases without limit, $[x \quad y \quad 1] = [1 \quad 0 \quad 0]$ which as we now know is a point at infinity on the x-axis. So the parabola has, as it should, two intersections with this axis, one at the origin, and one at infinity. Quadratic forms always have two solutions when intersected by lines. Both are real, or both are complex. These are the "roots" of the associated equations.

But now how about our hyperbola?
We can answer this question in the most straighforward way.

We have the parametric equation

$$[hx \quad hy \quad h] = [u^2 \quad u \quad 1] \begin{bmatrix} 1 & 1 & -1 \\ 0 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

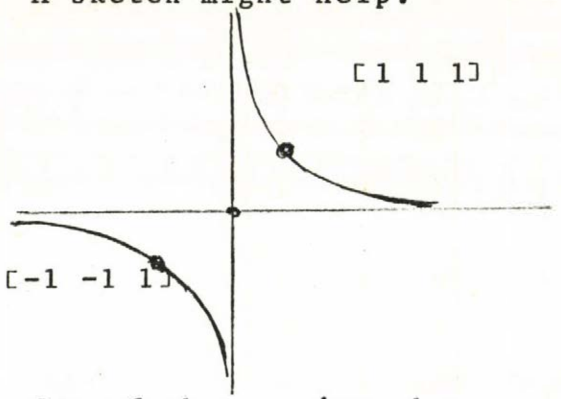
Now when $u = \infty$, this is

$$[hx \quad hy \quad h] = [1 \quad 0 \quad 0] \begin{bmatrix} 1 & 1 & -1 \\ 0 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix} = [1 \;\; 1 \;\; -1]$$

But according to the rules, the vector

$[1 \quad 1 \quad -1] \Rightarrow [-1 \quad -1 \quad 1]$, a point at $x = -1$, $y = -1$.

A sketch might help:

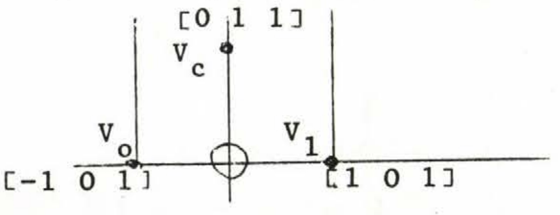


This indicates that the two "branches" of the hyperbola really belong to one curve, which passes through the two points at infinity.

One of these points happens when the parameter $u = 0$. The other occurs when $u = 1$. Notice that when $u = \infty$, the generated $[x \quad y \quad 1]$ point is a local one, at $x = -1$, $y = -1$.

- . - . - . -

Now we try another example.



Let $\begin{bmatrix} v_0 \\ v_T \\ v_1 \\ v_c \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

The reader can interpret v_o, v_1 and v_c easily, but $V_T = [0 \quad 1 \quad 0]$ represents the point at infinity on the Y-axis. (It will turn out that these four points yield a circle.)

The primitive conic is still $[x \quad y \quad 1] = [u^2 \quad u \quad 1]$, or $r = [u^2 \quad u \quad 1]$

We have already found that

$$r_c \begin{bmatrix} r_o \\ r_T \\ r_1 \end{bmatrix}^{-1} = [1 \quad 2 \quad 1], \text{ the } [a \quad b \quad c] \text{ vector.}$$

We need

$$v_c \begin{bmatrix} v_0 \\ v_T \\ v_1 \end{bmatrix}^{-1} = [0 \quad 1 \quad 1] \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^{-1}$$

Now

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

What we call the vector $[A \quad B \quad C]$ is now

$$v_c \begin{bmatrix} v_0 \\ v_T \\ v_1 \end{bmatrix}^{-1} = [0 \quad 1 \quad 1] \begin{bmatrix} -1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} = [1 \quad 2 \quad 1].$$

This yields the unexpected result,

$$[h_o \quad 2h_T \quad h_1] = [1 \quad 2 \quad 1], \quad \text{or } [h_o \quad h_T \quad h_1] = [1 \quad 1 \quad 1].$$

Substituting in the T matrix equation the h_i diagonal

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

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

Finally, the parametric equation for the image curve is

$$[hx \quad hy \quad h] = [u^2 \quad u \quad 1] \begin{bmatrix} 0 & -2 & 2 \\ 2 & 2 & -2 \\ -1 & 0 & 1 \end{bmatrix}$$

Now we shall see what kind of curve this represents.

We have $\qquad C = T^{-1}\, k\, T^{-1\,T}$

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

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

$$= \frac{1}{4} \begin{bmatrix} -1 & 1 & 1 \\ -1 & -1 & -1 \\ -1 & 1 & 1 \end{bmatrix}.$$

The quadratic form is (neglecting the 1/4 factor)

$$[x \quad y \quad 1] \begin{bmatrix} -1 & 1 & 1 \\ -1 & -1 & -1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$= [x \quad y \quad 1] \begin{bmatrix} (-x & +y & +1) \\ (-x & -y & -1) \\ (-x & +y & +1) \end{bmatrix} = \begin{matrix} -x^2 & +xy & +x \\ -xy & -y^2 & -y \\ -x & +y & +1 \end{matrix} = -x^2 - y^2 + 1 = 0.$$

This, by obvious rearrangement, is

$x^2 + y^2 = 1$, the equation for a unit circle centered at the origin.

We have now seen the primitive parabola $[x \;\; y \;\; 1] = [u^2 \;\; u \;\; 1]$ transformed into a hyperbola, and into a circle, by purely algebraic processes of projection. The matrix C arises from the product $(T^{-1}k\, T^{-1T},)$ and we would like to know where the little singular matrix k comes from.

For any C matrix,

$$[x \quad y \quad 1]\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}\cdot\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \quad \text{is the}$$

quadratic form. Performing the multiplications,

$$[x \quad y \quad 1] \quad \begin{bmatrix} (c_{11}x + c_{12}y + c_{13}) \\ (c_{21}x + c_{22}y + c_{23}) \\ (c_{31}x + c_{32}y + c_{33}) \end{bmatrix} = 0$$

$$\begin{array}{llll} c_{11}x^2 & + c_{12}xy & + c_{13}x & \\ c_{21}x & + c_{22}y & + c_{23}y & \\ c_{31}x & + c_{32}y & + c_{33} & = 0, \end{array}$$

and collecting like terms in x and y, this is

$$c_{11}x^2 + (c_{12} + c_{21})\, xy + c_{22}y^2 + (c_{13} + c_{31})\, x$$

$$+ (c_{23} + c_{32})\, y + c_{33} = 0.$$

This is the classic form

$$Px^2 + Qxy + Ry^2 + Sx + Ty + U = 0.$$

For our primitive parabola,

$[x \quad y \quad 1] = [u^2 \quad u \quad 1]$, we have, as we have seen, that $x = y^2$, or in proper canonical form, $x-y^2 = 0$.

When we substitute appropriate coefficients for P, Q, R, S, and T, we have

$$0.x^2 + 0.xy - y^2 + x + 0.y + 0 = 0$$

But
$$\begin{aligned}
P &= c_{11} = 0 \\
Q &= (c_{12} + c_{21}) = 0 \qquad\qquad c_{12} = -c_{21} \\
R &= c_{22} = -1 \\
S &= c_{13} + c_{31} = -1 \\
T &= (c_{23} + c_{33}) = 0 \qquad\qquad c_{32} = -c_{23} \\
U &= c_{33} = 0.
\end{aligned}$$

The resulting C matrix now looks like this:

$$\begin{bmatrix} 0 & c_{12} & c_{13} \\ -c_{12} & -1 & c_{23} \\ (1-c_{13}) & -c_{23} & 0 \end{bmatrix}.$$

In this matrix, there are the numbers c_{12} c_{13} and c_{23}. These numbers are undefined. Accordingly, we replace them with the simplest arbitrarily chosen set of three numbers we know.

Set $c_{13} = 1$, and c_{12} and c_{23} each equal to zero. The result is the matrix $\begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and this is only one of an infinity of matrices that will accomplish our purpose. We have called it the "k" matrix. Even though, it is doubly singular (or singularly-singular, as we might frivolously call it) it does what it's intended to do.

(This might suggest that even "singularly singular" matrices are not trivialities.)

By other choices of the numbers c_{12}, c_{13}, and c_{23} there result other matrices which have the same property as the simple k matrix we have used. Our choice was influenced by a yearning for simplicity.

Velocity, Accelaration, and Rate-of-change of Acceleration Vectors

$hv = [u^2 \quad u \quad 1] A$ is the parametric vector equation for a conic, and the Cartesian coordinates for the moving point are given by

$$v = \frac{hv}{h}, \qquad h \neq 0.$$

A is the constant matrix of coefficients. The first three derivations of hv are

$$\begin{aligned}
(hv)' &= h'v + hv' \\
(hv)'' &= h''v + h'v' + h'v' + hv'' \\
&= h''v + 2h'v' + hv'' \\
(hv)''' &= h'''v + h''v' + 2h''v' + 2h'v'' + h'v'' + hv''' \\
&= h'''v + 3h''v' + 3h'v'' + hv'''.
\end{aligned}$$

From these, we find

$$\begin{aligned}
hv' &= (hv)' - h'v \\
hv'' &= (hv)'' - h''v - 2h'v' \\
hv''' &= (hv)''' - h'''v - 3h''v' - 3h'v''
\end{aligned}$$

and of course

$$v' = \frac{hv'}{h} \qquad \text{the velocity of the moving point}$$

$$v'' = \frac{hv''}{h} \qquad \text{the acceleration}$$

$$v''' = \frac{hv'''}{h} \qquad \text{the rate-of-change of acceleration.}$$

Now $h = [u^2 \quad u \quad 1] \, A \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ a scalar from the third column of the A matrix.

Similarly,

$$h' = [2u \quad 1 \quad 0] \quad A \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$h'' = [\,2 \quad 0 \quad 0] \quad A \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$h''' = [\,0 \quad 0 \quad 0] \quad A \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \equiv 0 \text{ for all values of } u.$$

Also,

$$(hv)' = [2u \quad 1 \quad 0] \quad A$$

$$(hv)'' = [2 \quad 0 \quad 0] \quad A$$

$$(hv)''' = [0 \quad 0 \quad 0] \quad A \equiv [\,0 \quad 0 \quad 0\,], \text{a null vector}$$

for all u.

For the circle

$$A = \begin{bmatrix} 0 & -2 & 2 \\ 2 & 2 & -2 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad A \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}.$$

$$(hv)' = [2u \quad 1] \begin{bmatrix} 0 & -2 & 2 \\ 2 & 2 & -2 \end{bmatrix}$$

$$h' = [2u \quad 1] \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

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

$$h = [u^2 \quad u \quad 1] \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

In particular, when $u = \frac{1}{2}$,

$$(hv)' = [2 \quad 0 \quad 0]$$

$$h' = 0$$

$$hv = [1/4 \quad 1/2 \quad 1]\begin{bmatrix} 0 & -2 & 2 \\ 2 & 2 & -2 \\ -1 & 0 & 1 \end{bmatrix} = [0 \quad 1/2 \quad 1/2]$$

$$h = \frac{1}{2} \qquad v = [0 \quad 1 \quad 1]$$

Then

$$hv' = (hv)' - h'v = [2 \quad 0 \quad 0]$$

$$v' = \frac{hv'}{h} = [4 \quad 0 \quad 0]$$

For the acceleration vector,

$$(hv)'' = [2 \quad 0 \quad 0]\begin{bmatrix} 0 & -2 & 2 \\ 2 & 2 & -2 \\ -1 & 0 & 1 \end{bmatrix} = [0 \quad -4 \quad 4]$$

$h'' = 4$
$h' = 0$
$v = [0 \quad 1 \quad 1]$
$v' = [4 \quad 0 \quad 0]$

Then
$$\begin{aligned} hv'' &= (hv)'' - h''v - 2h'v' \\ &= [0 \quad -4 \quad 4] - [0 \quad 4 \quad 4] \\ &= [0 \quad -8 \quad 0] \end{aligned}$$
and
$$v'' = \frac{hv''}{h} = [0 \quad -16 \quad 0] .$$

– . – . – . –

For the third derivative vector, $(hv)''' \equiv [0\ 0\ 0]$ and $h''' \equiv 0$

$$hv''' = -3\,h''v' - 3h'v''$$

$$= -12\,[4\ 0\ 0], \quad \text{since} \quad h' = 0 \text{ in this case.}$$

$$v''' = \frac{hv'''}{h} = -24[4\ 0\ 0] = [-96\ 0\ 0].$$

- . - . - . -

THREE-DIMENSIONAL PROJECTIVE TRANSFORMATIONS

All of the foregoing involved manipulations with vectors, and in homogeneous coordinates these vectors consisted of three components, describing points and lines in a two-dimensional space.

It seems now natural to consider vectors of four components, describing points and <u>planes</u> in a three-dimensional space.

The analogy is not a mere semantic device.
The product

$$\begin{bmatrix} hx & hy & hz & h \end{bmatrix} \begin{bmatrix} \alpha A \\ \alpha B \\ \alpha C \\ \alpha D \end{bmatrix} = 0$$

expands:

$$h\alpha \quad Ax + h\alpha \quad By + h\alpha \quad Cz + h\alpha D = 0$$

which is recognized as the traditional equation of a plane.
The quantities h and α are entirely arbitrary.

The matrix $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ describes, by rows,

the homogeneous coordinates of
a point at infinity on the x axis,
a point at infinity on the y axis,
a point at infinity on the z axis,
and the origin of coordinates of the three-dimensional space.

This matrix of vectors is sometimes called the "simplex" of the space. Algebraically, of course, it is just the identity matrix of the system.

Suppose there is some transformation matrix, T, such that the object-space image-space transformation is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} T = \begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{bmatrix}$$

The four vectors $[a_i \quad b_i \quad c_i \quad d_i]$ are the images of their corresponding points in the object space. They may or may not lie, themselves, at infinity, depending upon the value of the d_i. Observe the striking fact that even the origin of coordinates, $[0 \quad 0 \quad 0 \quad 1]$ may quite possibly become a point at infinity in the transformed space.

Such a case is <u>not</u> a degeneracy of the transformation. It is entirely normal.

The coordinates of the four transformed points in the image space, taken together in the matrix, <u>are</u> the components of the transformation matrix T. This gives an interesting and illuminating interpretation of the T matrix.

As we have done before in the case of point-line coindidence in the plane, we examine point-plane coincidence.

Abbreviate the notation, and write

$p = [hx \quad hy \quad hz \quad h]$ for the point vector, and

$$\pi = \begin{bmatrix} \alpha A \\ \alpha B \\ \alpha C \\ \alpha D \end{bmatrix}$$ for the plane vector.

Then $p\,\pi = 0$ describes the combined position of point and plane. Now introduce the product of T and its adjoint, T*, such that

$TT^* = [\text{diag}]$ a diagonal matrix, all of whose elemenst are identical. They may all be zero, without invalidating what follows.

Then

$$p^{TT^*}\pi = 0 \qquad \text{certainly.}$$

But $\quad p^{T} = p^*,\quad$ a point transformation.

Similarly, we can regard $T^*\pi$ as a plane transformation, yielding the new plane π^*, so that

$$p^*\pi^* = 0.$$

This describes the image of the point-plane coincidence after the transformation of the space. It is of course an invariant of the transformation.

As in the two-dimensional case, we can partition the T matrix,

$$\left[\begin{array}{ccc|c} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ \hline a_4 & b_4 & c_4 & d_4 \end{array}\right] = T$$

and consider the significance of each partition.

The most interesting partition consists of the elements d_i in the fourth column.

Consider

$$T = \left[\begin{array}{ccc|c} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & d_3 \\ \hline 0 & 0 & 0 & d_4 \end{array}\right]$$

where no affine transformation takes place (the upper left-hand partition is just the identity matrix.) and where no displacement of the coordinate system happens (the lower left-hand partition is null.)

Now we form the adjoint of this matrix.
It is

$$T^{*} = \begin{bmatrix} d_4 & 0 & 0 & -d_1 \\ 0 & d_4 & 0 & -d_2 \\ 0 & 0 & d_4 & -d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

as can be easily verified, as follows.

$$\underset{T}{\begin{bmatrix} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & d_3 \\ 0 & 0 & 0 & d_4 \end{bmatrix}} \underset{T^{*}}{\begin{bmatrix} d_4 & 0 & 0 & -d_1 \\ 0 & d_4 & 0 & -d_2 \\ 0 & 0 & d_4 & -d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}} = \begin{bmatrix} d_4 & 0 & 0 & 0 \\ 0 & d_4 & 0 & 0 \\ 0 & 0 & d_4 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix}$$

If d_4 is not zero, the inverse of T is simply

$$T^{-1} = \frac{T^{*}}{d_4} .$$

But if $d_4 = 0$, the matrix T has no inverse, even though it <u>does</u> possess an adjoint, invariably.

The transformation of a plane is $T^{*}\pi$, which is

$$\begin{bmatrix} d_4 & 0 & 0 & -d_1 \\ 0 & d_4 & 0 & -d_2 \\ 0 & 0 & d_4 & -d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} Ad_4 - Dd_1 \\ Bd_4 - Dd_2 \\ Cd_4 - Dd_3 \\ D \end{bmatrix}$$

This shows that every plane in the object space has its unique image, so long as $d_4 \neq 0$.

But if $d_4 = 0$, all planes in the object space have the same image,

$$\begin{bmatrix} -Dd_1 \\ -Dd_2 \\ -Dd_3 \\ D \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} -d_1 \\ -d_2 \\ -d_3 \\ 1 \end{bmatrix}$$

When $d_4 = 0$, we obtain a plane image of a three-dimensional object space. But when $d_4 \neq 0$, we obtain what has been called a "relief perspective" of the object space. A crude (and imprecise) notion of relief perspectives is given,for example, by the images of scenes and prominent persons embossed on coins.

It is, paradoxically, false to think that the pictures made by a photographic device (a camera) are plane images. They are really two-dimensional slices of a three-dimensional image formed optically inside the camera. The operation of focussing the camera serves to choose the appropriate slice of this intrinsically three-dimensional image.

- . - . - . -

FOUR-DIMENSIONAL TRANSFORMATIONS, ORDINARY COORDINATES

A point in four dimensions is represented by the four-component vector [x y z w]. These are "ordinary" cartesian coordinates, not projective ones, and refer to the four mutually orthogonal unit vectors of the space. The simplex

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

describes these four orthogonal unit vectors.

The origin of coordinates of the system is the local point, [0 0 0 0].

Any 4x4 matrix describes an affine transformation of the space, and its rows describe the images of the four "unit" points of the simplex.

A special affine transformation of interest is the rigid rotation of the system with its embedded configurations.

If [a b c d] is the rotated image of one of the unit vectors, then its length must still be a unit, or

$$[a \quad b \quad c \quad d] \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = 1$$ (the square of the length.)

If $[a_1 \; b_1 \; c_1 \; d_1]$ and $[a_2 \; b_2 \; c_2 \; d_2]$ are the images of two of the orthogonal unit vectors, then after the rotation, we want them to remain orthogonal; this implies

$$[a_1 \quad b_1 \quad c_1 \quad d_1] \begin{bmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{bmatrix} = 0.$$ Similarly for the others.

We can summarize these two requirements:

$$\begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This says that if a matrix T satisfies

$T\,T^T = I$, the identity matrix, then

$T^T = T^{-1}$.

The transpose of the matrix is also the inverse of the matrix. Such matrices are called "ortho-normal" systems. They are strictly length and angle preserving systems, and are consequently shape-invariant.

We can make two-dimensional pictures of objects embedded in such coordinate systems. We need only multiply the ortho-normal system matrix by a "projector" matrix that "selects" the desired picture coordinates and "rejects" the others. This is precisely what we do when we make an "orthographic" or di-metric or tri-metric or isometric or "oblique" drawing of an object in 3-space.

For instance,

$$\begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & & \\ a_2 & b_2 & & 0 \\ a_3 & b_3 & & \\ a_4 & b_4 & & \end{bmatrix}$$

These are just the x and y coordinates of the points in the 4-dimensional object. Of course the z and w coordinates are missing, as they should be. The resulting figure is a picture of the object.

Similarly, of course, we could construct a three-dimensional image of the four-dimensional object.

- . - . - . -

CONSTRUCTION OF A 4x4 ORTHONORMAL MATRIX

Choose four arbitrary quantities for the top row.
Choose three quantities for the second row,
two quantities for the third row,
and one arbitrary quantity for the fourth row.

$$\begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & p \\ a_3 & b_3 & q & r \\ a_4 & s & t & u \end{bmatrix}$$

In this example, the quantities p, q, r, s, t, u are to be found so as to satisfy the orthogonality conditions.

Thus $\begin{bmatrix} a_1 & b_1 & c_1 & d_1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ p \end{bmatrix} = 0$ is sufficient to determine p.

Now $\begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & p \end{bmatrix} \begin{bmatrix} a \\ b \\ q \\ r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ determines q and r,

and finally

$\begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & p \\ a_3 & b_3 & q & r \end{bmatrix} \begin{bmatrix} a_4 \\ s \\ t \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ determines s t and u.

The resulting matrix is orthogonal, but not yet orthonormal.

The product of this matrix and its transponse yields a diagonal matrix,

$$\begin{bmatrix} A & 0 & 0 & 0 \\ 0 & B & 0 & 0 \\ 0 & 0 & C & 0 \\ 0 & 0 & 0 & D \end{bmatrix} .$$ We now need only

to multiply the orthogonal matrix by the appropriate "scaling matrix", which is

$$\begin{bmatrix} \frac{1}{\sqrt{A}} & & & \\ & \frac{1}{\sqrt{B}} & & \\ & & \frac{1}{\sqrt{C}} & \\ & & & \frac{1}{\sqrt{D}} \end{bmatrix}$$

to obtain the desired orthonormal matrix.