Original Lecture #2: 5 October 1993
Topics: Homogeneous Coordinates
Scribe: Arturo Puente*

1 Why homogeneous coordinates?

Some of the reasons for working with homogeneous coordinates were mentioned by Stolfi [Stolfi 91]:

- Simpler formulas
- Fewer special cases
- Unification and extension of concepts
- Duality

Let's start with an example (see Figure 1). The two affine equations

$$2X - 3Y = 5$$

$$4X - 6Y = 2$$

represent two parallel lines, which do not intersect in the affine plane. But, if homogenize both equations, we obtain

$$2x - 3y = 5w$$

$$4x - 6y = 2w$$

and this system has the following family of solutions

$$w = 0$$

$$x = 3c$$

$$y = 2c$$

for any real number c. This solution represents the point at infinity $[0;3,2]$ in the direction of the vector $(3,2)$.

If the solution of the homogenized system is of the kind

$$w = ac$$

$$x = bc$$

*these notes were originally scribed when the course was offered in 1991

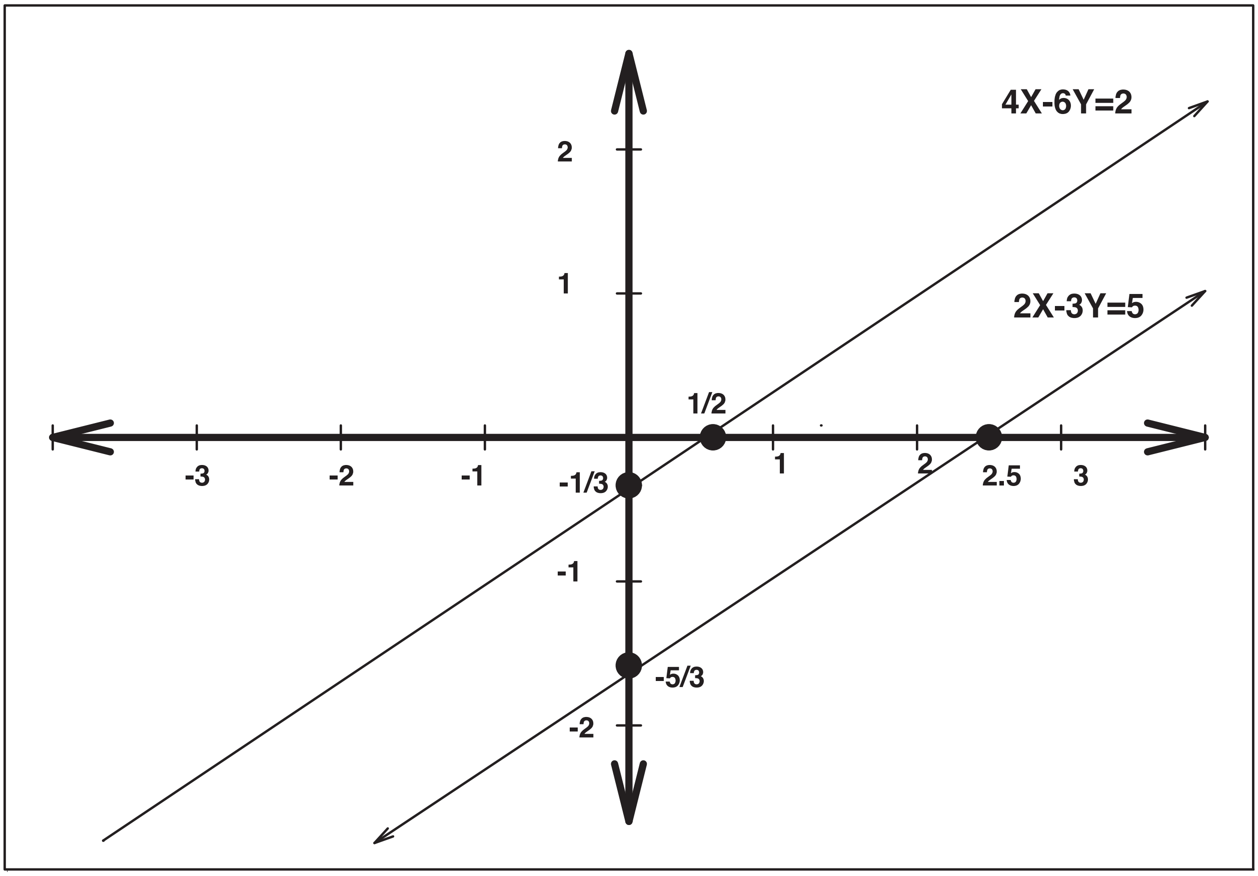

Figure 1: Example 1

$$y = dc$$

with $a \neq 0$, we can choose $c = 1/a$ and get $w = 1$. The resulting solution $[w;x,y] = [1;b/a,d/a]$ of the homogenized system corresponds to the point $(b/a,d/a)$ in the affine plane. In this course we use column vectors to represent points. To remind ourselves of that, we use square brackets to represent points and angled brackets to represent lines (which are row vectors).

Another example: Let's consider the following pair of equations:

$$X + Y = 2$$

$$X - Y = 0$$

We can homogenize the equations to obtain:

$$x + y = 2w$$

$$x - y = 0w$$

The solution of this system of equations is

$$w = c$$

$$x = c$$

$$y = c$$

for any real number c. If we want to make $w = 1$, we can choose $c = 1$, getting $[w;x,y] = [1;1,1]$. When the affine system has a solution, the homogenized, linear system has a solution with $w = 1$. In this course, we will often use homogenization to convert affine equations and system into linear equations and systems.

$$\text{affine equations} \quad \overset{\text{homogenize}}{\longmapsto} \quad \text{linear equations.}$$

We are now going to study how homogeneous coordinates help us in our work with algebraic geometry (see Figure 2).

Let's consider the intersection between the curve $XY = 1$ and several other curves. The first case is the intersection with the line $X + Y = \frac{5}{2}$ we have:

$$Y = \frac{5}{2} - X$$

$$X(\frac{5}{2} - X) = 1$$

$$0 = X^2 - \frac{5}{2}X + 1$$

$$0 = (X-2)(X-1/2)$$

In this case we have two distinct, real roots.

The second case is the intersection with the line $X + Y = 2$, we have:

$$X + Y = 2$$

$$Y = 2 - X$$

$$X(2-X) = 1$$

$$0 = X^2 - 2X + 1$$

$$0 = (X-1)(X-1)$$

In this case we have two equal real roots, that is, a double root.

The third case is the intersection with the line $X + Y = 0$, we have:

$$X + Y = 0$$

$$Y = -X$$

$$X(-X) = 1$$

$$0 = X^2 + 1$$

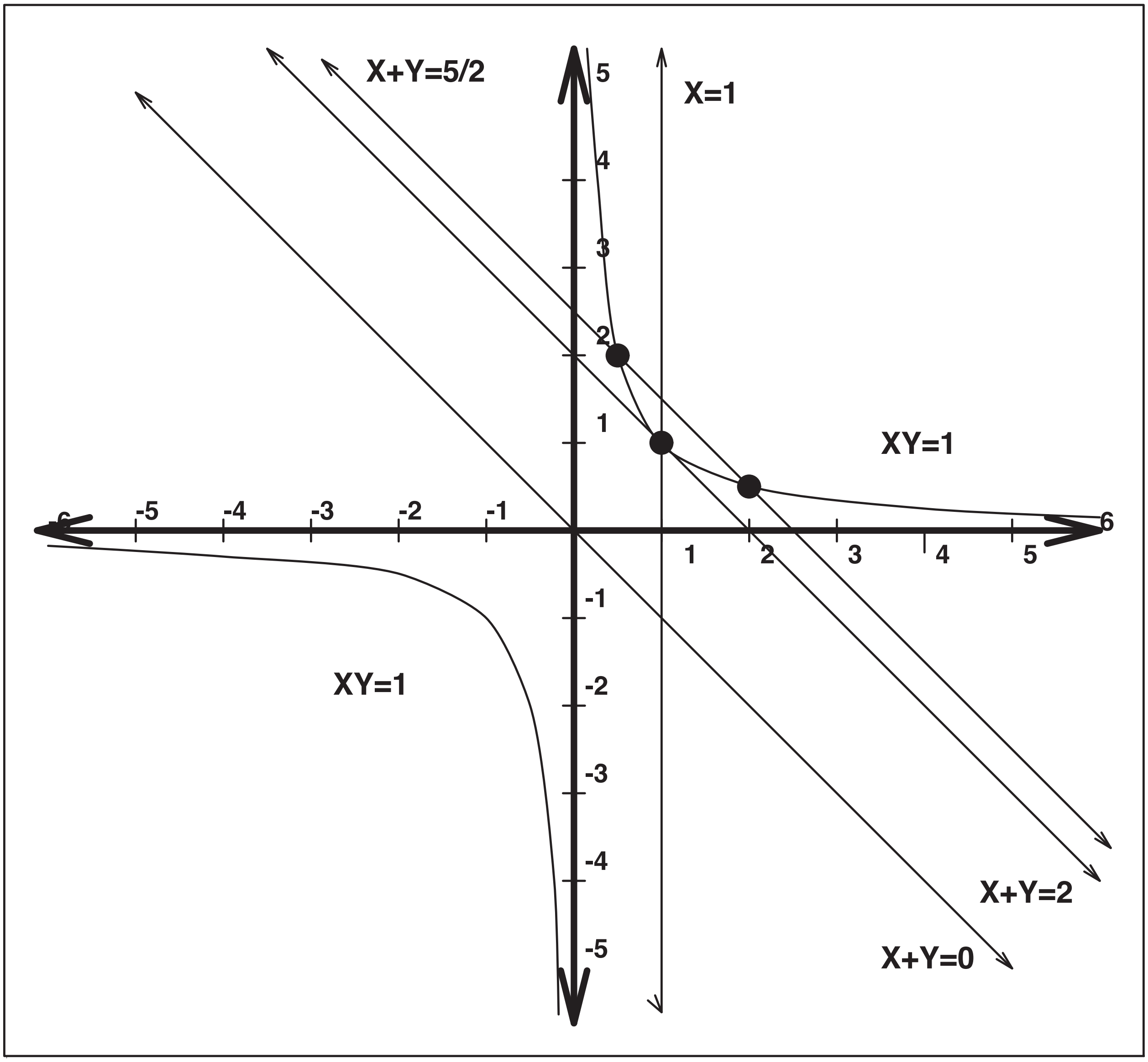

Figure 2: Example 2

$$0 = (X+i)(X-i)$$

In this case we have two distinct, complex roots.

The fourth case is the intersection with the line $X = 0$. If we proceed naively, we have

$$X = 0w$$

$$XY = 1$$

$$0Y = 1$$

$$0 = 1$$

but that doesn't make sense. Instead, we first homogenize the equations, getting:

$$x = 0$$

$$xy = w^2$$

$$0 = w^2$$

$$0 = (w-0)(w-0)$$

This implies that $w = 0$ is a double root. When $w = 0$, we also have $x = 0$ by assumption. But we can't also have $y = 0$, or we would be in the illegal case of having all three homogeneous coordinates equal to 0. So we must have $y \neq 0$, say $y = 1$. Thus, the hyperbola intersects the line $X = 0$ twice at the point $[w;x,y] = [0;0,1]$, the point at infinity in the vertical direction. In fact, the line $X = 0$ is the tangent line to the hyperbola $XY = 1$ at the point $[0;0,1]$.

The fifth case is the intersection with the line $X = 1$. In this case, we once again homogenize the equations:

$$x = w$$

$$xy = w^2$$

$$wy = w^2$$

$$0 = w^2 - wy$$

$$0 = w(w-y)$$

$$0 = (w-0)(w-y)$$

So we have two points of intersection, one with $w = 0$ and the other with $w = y$. When $w = 0$, we have $x = 0$ also, so, as in the fourth case, we must have $y \neq 0$ and the resulting point is $[w,x,y] = [0;0,1]$, the point at infinity in the vertical direction. When $w = y$, we also have $x = w$, so the resulting point is $[w,x,y] = [1;1,1]$, the finite point whose Cartesian coordinates are $(1,1)$.

This example shows how problems that have no solution in the affine plane can have solutions in the projective plane, which we can locate by using homogeneous coordinates. In fact, if we count multiple roots properly and count complex roots and count roots at infinity, then any two plane curves, one of degree m and the other of degree n, either intersect in precisely mn points or in infinitely many points (Bézout's Theorem). Thus, the projective approach provides the right number of solutions for each problem and all those solutions can be expressed in the same form, avoiding any need for exceptions or special cases. As a result, algorithms can be coded with fewer separate cases and things are generally more elegant.

We convert a point expressed in Cartesian coordinates to homogeneous coordinates by adding $w = 1$ as the additional, weight coordinate. In the plane, we have

$$(X,Y) \text{ Cartesian} \Longrightarrow [1;X,Y] \text{ Homogeneous}$$

and in space, we have

$$(X,Y,Z) \text{ Cartesian} \Longrightarrow [1;X,Y,Z] \text{ Homogeneous.}$$

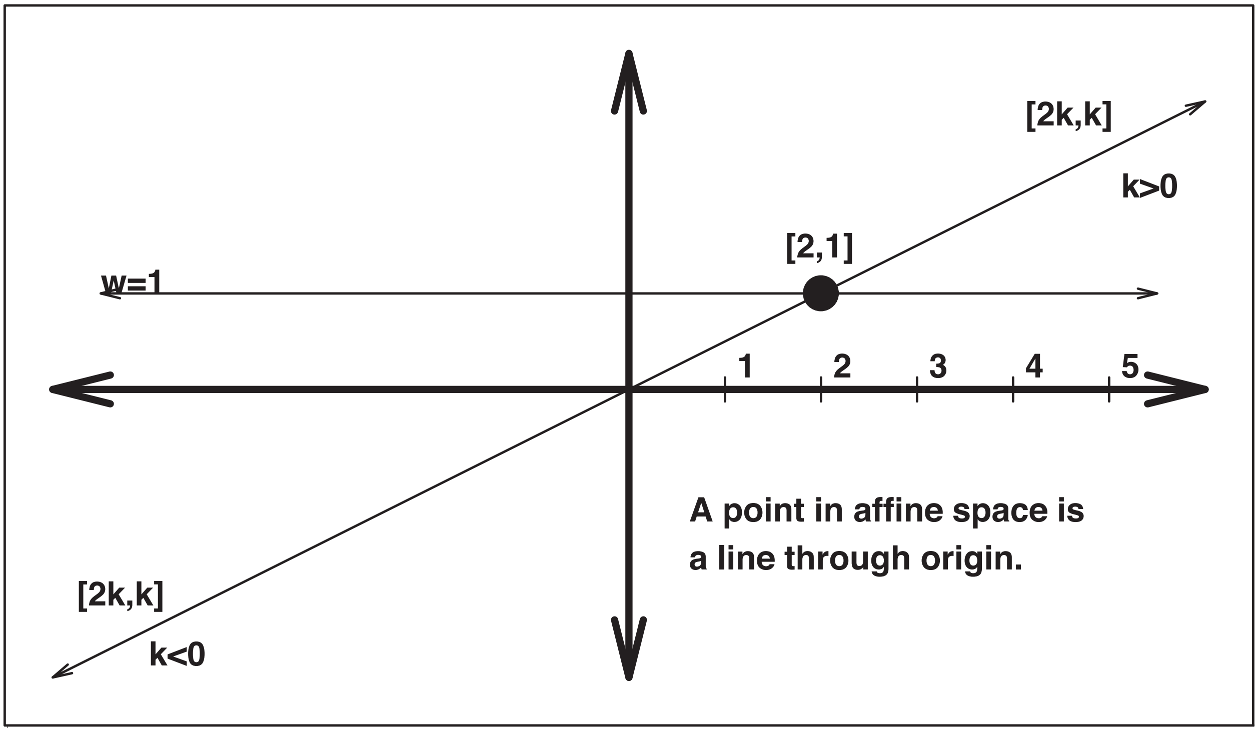

Figure 3: Line in site space that represents the point $(2) = [2,1]$.

Of course, we can also multiply all of the homogeneous coordinates by any nonzero scalar without changing the corresponding point. So it is equally valid, say in the plane, to take

$$(X,Y) \text{ Cartesian} \Longrightarrow [w;Xw,Yw] \text{ Homogeneous}$$

for any nonzero w. In the special case of $w = 1$, the resulting homogeneous coordinates are called *normalized*.

We convert from homogeneous coordinates to Cartesian coordinates by the following rule:

$$[w;x,y] \text{ and } w \neq 0 \Longrightarrow (X,Y), \text{ where } X = \frac{x}{w} \text{ and } Y = \frac{y}{w}$$

for 2-dimensions, and

$$[w;x,y,z] \text{ and } w \neq 0 \Longrightarrow (X,Y,Z), \text{ where } X = \frac{x}{w}, Y = \frac{y}{w}, \text{ and } Z = \frac{z}{w}$$

for 3-dimensions (see Figure 3 for a picture of how this projection works when going from a point in projective 1-space, written in homogeneous coordinates, to the corresponding point in affine 1-space, written in Cartesian coordinates).

The homogeneous points with w coordinate equal to 0 are points at infinity. For example, $[0;1,3]$ is the point at infinity with direction $(1,3)$ and $[0;2,4,5]$ is the point at infinity with direction $(2,4,5)$.

1.1 Equation of a line in homogeneous coordinates

The equation of a line in Cartesian coordinates is:

$$Y = mX + b$$

where m is the *slope* and b is the *Y-intercept*, that is, the value of Y when $X = 0$. In the case of homogeneous coordinates, we associate with a line three *homogeneous coefficients*. These coefficients are calculated so that

$$\langle a; b, c\rangle = \{[w; x, y] \mid aw + bx + cy = 0\}$$

Given a line equation $Y = mX + b$ in slope-intercept form, we can rewrite that equation as

$$b + mX - Y = 0$$

so the homogeneous coefficients of that line are

$$\langle a; b, c\rangle = \langle b; m, -1\rangle.$$

The homogeneous coefficients of the line at infinity are $\langle 1; 0, 0\rangle$.

1.2 Homogeneous line defined by two points

Suppose that we have two points defined in homogeneous coordinates

$$P_1 = [w_1; x_1, y_1]$$

$$P_2 = [w_2; x_2, y_2]$$

and that we want the homogeneous coefficients of the line l that joins P_1 to P_2. Consider a third point

$$P = [w; x, y]$$

Let us characterize the condition under which P lies on l. If P_1, P_2, and P are collinear, then the triangle defined by the three points has zero area. Using the determinant formula for the area of a triangle, this condition can be rewritten as,

$$\begin{vmatrix} w & x & y \\ w_1 & x_1 & y_1 \\ w_2 & x_2 & y_2 \end{vmatrix} = 0$$

This condition exactly characterizes the line joining P_1 and P_2 and therefore expanding the above determinant gives the equation of the line. Hence the homogeneous coordinates of the line l are,

$$\langle x_1 y_2 - y_1 x_2; y_1 w_2 - w_1 y_2, w_1 x_2 - x_1 w_2$$

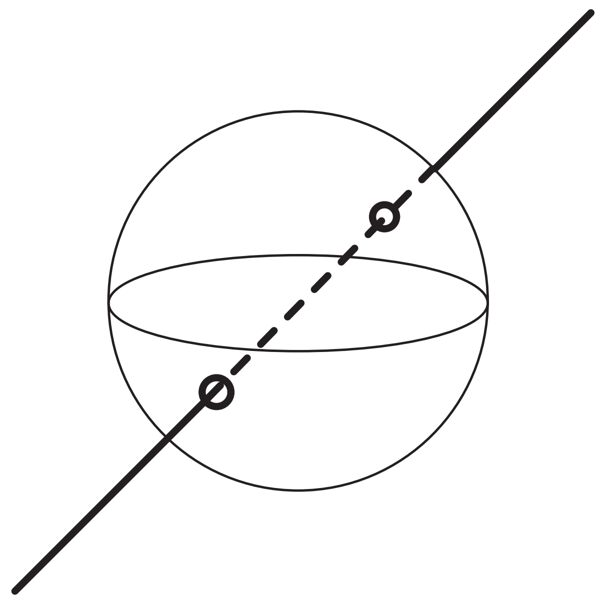

Figure 4: The spherical model of the projective plane $\mathbf{P_2}$

Similarly, given the homogeneous coefficients for two lines, we can compute the homogeneous coordinates of the intersection point. Suppose

$$l_1 = \langle a_1; b_1, c_1 \rangle$$

$$l_2 = \langle a_2; b_2, c_2 \rangle$$

A third line $l = \langle a; b, c \rangle$ is concurrent with l_1 and l_2 if and only if the determinant

$$\begin{vmatrix} a & b & c \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0$$

The homogeneous coordinates $(w; x, y)$ of the point of intersection are given by expanding the above determinant. Hence, the homogeneous coordinates of the point are,

$$(b_1c_2 - b_2c_1; a_2c_1 - a_1c_2, a_1b_2 - a_2b_1)$$

In projective space, we work with lines through the origin instead of points, and planes through the origin instead of lines. But, to keep the terminology simple, we continue using the words

- *point* for a line through the origin of site space.
- *line* for a plane through the origin of site space.

At this point Prof. Guibas talked abut the different models of the projective plane, especially the straight model and the spherical model, and also talked about the topology of the projective plane (see Figure 4).

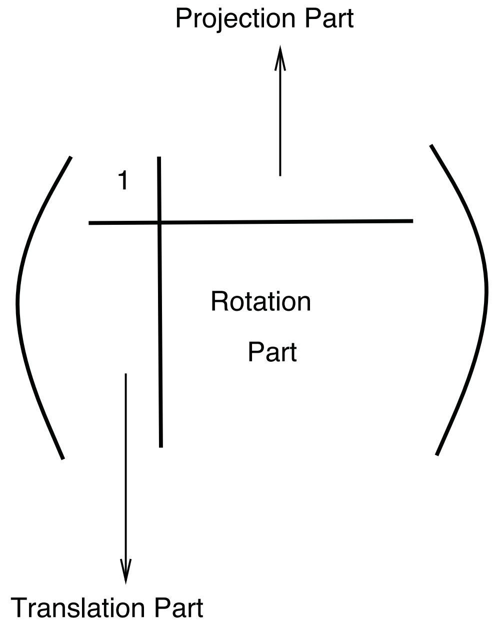

Figure 5: Components of a projective transformation matrix

2 Matrices for linear, affine, and projective transformations.

An affine map of the plane has 6 degrees of freedom. One way to specify such a map is by giving three points in the domain (non-collinear) and specifying where each of those three points goes in the range. Similarly, an affine map of 3-space can take any four points (non-coplanar) to any four points.

Projective maps have more freedom. A projective map of the plane can take any four points, no three collinear, to any other four points, no three collinear. A projective map of space can take any five points, no four coplanar, to any other five points, no four coplanar.

Another way to think of a projective map, or *projectivity*, is as the composition of some number of perspectivities, where a *perspectivity* is the map—say, from one plane to another—that results from looking from some vantage point through one plane at the other plane. The composition of two perspectivities is not necessarily a perspectivity. But any composition of perspectivities is a projectivity.

Both affine and projective maps of the plane can be represented as 3-by-3 matrices, and both affine and projective maps of 3-space can be represented as 4-by-4 matrices. The matrix that represents a projective map can be divided into several submatrices, which are responsible for the different components of the transformation. We can identify three major parts, as shown in Figure 5: the rotation component, the translation component, and the projection component. The matrix represents an affine map precisely when the projection component is all zeros, that is, when the first row of the matrix is $(1,0,0,\ldots,0)$. (Figure 5 shows a 1 in the top-left corner of the matrix. If the matrix represents a projective map, we can multiply all of the matrix elements by any common, nonzero scalar without changing anything. Thus, as long as the

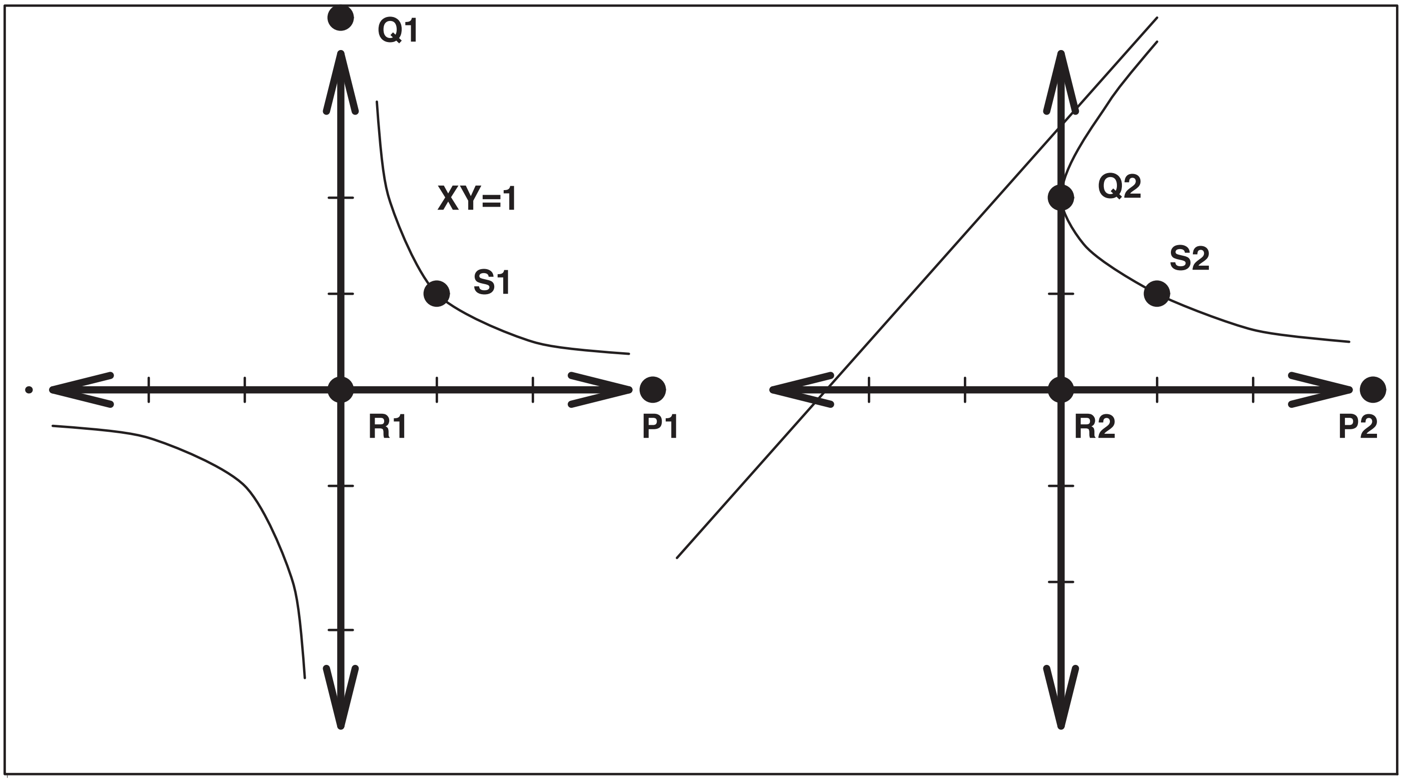

Figure 6: A projective map of the hyperbola $XY = 1$

top-left element isn't 0, we can normalize the matrix to make its top-left element be 1. When the matrix represents an affine map, we always normalize it in this way.)

3 An example of a projective map

Let's find a projective map that allows us to see the intersection point Q_1 of the hyperbola $XY = 1$ with the line $X = 0$. In particular, let's move the point $Q_1 = [0;0,1]$, which is the point at infinity in the vertical direction, to the finite point $Q_2 = [1;0,2]$. While we move Q_1 to Q_2, let's keep fixed the following three other points: the point $P_1 = [0;1,0]$, the point at infinity in the horizontal direction; the point $R_1 = [1;0,0]$, the origin; and the point $S_1 = [1;1,1]$. That is, let's make $P_2 = P_1$, $R_2 = R_1$, and $S_2 = S_1$. (See figure 6). Recall that a projective map of the plane can take any four points, no three collinear, to any other four points, no three collinear; so what we want to do makes sense.

All that we start out knowing is that our projective map is represented by some 3-by-3 matrix, say M:

$$M = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}$$

The condition $[M][P_1] = [P_2]$ tells us that

$$\begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} d \\ e \\ f \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

so we must have $d = 0$ and $f = 0$. Similarly, the condition $[M][R_1] = [R_2]$ tells us that

$$\begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

so we must have $b = 0$ and $c = 0$.

The condition $[M][Q_1] = [Q_2]$ is a bit more complicated:

$$\begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} g \\ h \\ i \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

tells us that $h = 0$ and that $i = 2g$. So far, we have deduced that the matrix M must have the form

$$M = \begin{bmatrix} a & 0 & g \\ 0 & e & 0 \\ 0 & 0 & 2g \end{bmatrix}.$$

But we still don't know the ratios $a : e : g$.

By determining these final ratios properly, we can arrange that $[M][S_1] = [S_2]$:

$$\begin{bmatrix} a & 0 & g \\ 0 & e & 0 \\ 0 & 0 & 2g \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} a+g \\ e \\ 2g \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

That is, we want $a + g = e = 2g$. One easy way to achieve this is to let $a = 1$, $e = 2$, and $g = 1$. The resulting projective map M is

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

This projectivity M maps Q_1 to Q_2 and R_1 to R_2, so it maps the Y-axis to itself. It turns out that it maps the hyperbola $XY = 1$ to a hyperbola with the lines $Y = 0$ and $Y = X + 4$ as its asymptotes. This new hyperbola is tangent to the Y-axis (the line $X = 0$) at the point Q_2.

References

[Foley 90] J.D. Foley, A. van Dam, S.K. Feiner, J.F. Hughes. "Computer Graphics. Principles and Practice," *Addison-Wesley Publishing Company*, Appendix: Mathematics for Computer Graphics: 1083-1112, 1990.

[Stolfi 91] J. Stolfi, "Oriented Projective Geometry. A framework for Geometric Computations," *Academic Press*, First Edition. Chapter 1. Projective geometry: 3-11, 1991