## 2. Projective Geometry in General

### 2.1. Some History

Projective geometry is by no means a new field of mathematics. Some of its very classic theorems cited in all basic textbooks (eg theorems of Pappos and Menelaos) are of Greek origin. It can be supposed that ancient Greek mathematicians knew much more about projective geometry; unfortunately, most of their work has been lost and only some indirect facts can be used to measure the exact amount of their knowledge.

It was in the period of the Renaissance that projective geometry gained a much greater importance. This was the result of the fact that artists at that time were greatly interested in creating realistic pictures; as opposed to medieval painters they wanted to understand exactly how a three dimensional object could be rendered on a two dimensional plane, that is a canvas. And this is basically what projective geometry is all about. At a time when it was (still) natural that artists would also work on "scientific problems" if they felt the necessity for it, Pietro della Francesca (1420-1492) or Albrecht Dürer (1471-1528) wrote down their ideas about the rules of projection; the book of Dürer ([Düre66]) is probably one of the first books ever written on the subject. It is also interesting to note that he produced a number of carvings in which practical techniques of how to make projections are presented in an artistic way.

A more concise mathematical investigation on projective geometry was started by G. Desargues (1593-1662). It was he who has introduced the notion of a point and a line at infinity. His work was followed by a number of other mathematicians (B. Pascal, L.N. Carnot, G. Monge and others) who discovered those theorems and facts about projective geometry which still form the basis of the theory today.

However, the exact role of projective geometry in the description of the surrounding world was for a long time somewhat fuzzy. Indeed, one has to understand that at that time geometry as well as the philosophy of nature in general was very much dominated by Euclidean geometry. In his monumental work *The Elements* ([Eukl75]), which was a kind of an encyclopaedia of Greek geometry, Euclid had created the first axiomatic system in history of mathematics; his work was so successful that up to the $19^{th}$ century everybody thought that Euclidean geometry was not only an efficient tool to describe nature but the surrounding world *was* Euclidean. This belief was even more strengthened by the fact that Newton's *Principia* was very much based on Euclidean geometry when describing the rules of mechanics. Very typically, the "Princeps Mathematicae", F. Gauss (1777-1855), who was probably one of the first mathematicians to realise that the truth might be different, did not dare to publish his results; he was afraid to be in conflict with all the great intellects of his time.

This firm belief in the overall nature of Euclidean geometry was tarnished by the independent works of J. Bólyai and N.I. Lobatchewsky. Indeed, these two mathematicians succeeded in creating a new geometry (which later received the name of *hyperbolic geometry*) which was fundamentally different from the Euclidean one. This geometry was the result of their investigations on Euclid's so called $5^{th}$ postulate, which said that if a point does not intersect a line, then there is *only one* line intersecting this point which is parallel to the given line. After centuries of unsuccessful attempts to *prove* this postulate out of the remaining axioms of Euclidean geometry, both Bólyai and Lobatchewsky created a new axiomatic system by taking all other Euclidean axioms and the *negation* of the $5^{th}$ postulate (that is that there exist *more than one* distinct parallel line intersecting the external point). This new geometry is very different in flavour: the sum of the internal angles of a triangle is not 180°, the "traditional" trigonometrical equations are no longer valid etc. It is, however, undecidable whether Euclidean geometry or the hyperbolic one is the adequate description of reality; in fact, both of them are only *models* and one could as well describe the whole surrounding world by using hyperbolic geometry instead of the Euclidean one.

Besides the technical nature of this result, the birth of hyperbolic geometry has shown that there might be a whole range of different "geometries", each of them modelling some particular aspect of reality. One may speak of multidimensional geometry (giving a geometrical structure to the set of vectors of higher dimensions), of complex geometry, of hyperbolic and elliptic geometries etc. Projective geometry turned out to be one of these different geometries, a useful tool in the description of some phenomena concerning projections, lines, planes and so forth.

These results (together with other advances in mathematics in the $19^{th}$ century) had also strengthened the need of a precise foundation of all mathematical fields, including projective geometry. This was performed by the thorough use of the axiomatic method and of set theory, which has become the fundamental basis of mathematics in our century and had been initiated largely by the *Erlanger Programm* of F. Klein (1849–1925) and by the *Grundlagen der Geometrie* of D. Hilbert (1862–1943). Projective geometry has also been reformulated in this way: there is a very precise set of axioms which defines projective geometry and the existence of this axiomatic approach gives also a very precise insight of how projective geometry is related to other fields of mathematics (specially Euclidean geometry). This system of axioms will be presented in a later section.

In the $20^{th}$ century traditional projective geometry has lost its momentum as a field of basic mathematical research; in this sense it might be considered as a "classical" theory$^{\dagger}$. It has by no means lost importance though, having given birth to a whole range of practical tools and methods used to make drafts and technical drawings throughout the world. It is the very aim of this present work to show that a more precise knowledge of projective geometry can also play a significant role in

$^{\dagger}$ Today's researches are more directed toward algebraic and combinatorical problems arising when investigating for example *finite* projective spaces; these problems have had, however, no relevance for computer graphics up to now.

giving new and perhaps more understandable approaches to the methods and algorithms used in computer graphics.

## 2.2. Notational Conventions

Before going further some notations are listed which will be used throughout the thesis.

Points will be denoted by capital Latin characters ($A$, $A'$, $A''$, $P$, $Q$ and also $A_i$, $B_j$ etc.) whereas small Latin characters ($a$, $a'$, $a''$, $n$, $l$ etc.) will denote lines. 2D subplanes of the Euclidean space will be denoted by capital Greek letters like $\Pi$, $\Pi'$, $\Pi''$, $\Phi$, $\Psi$. The symbol "$\wedge$" will be used to denote intersection; that is $a \wedge b$ will give the intersection of the lines $a$ and $b$ while $\Pi \wedge b$ will denote the intersection of the plane $\Pi$ and the line $b$. Symmetrically, the symbol "$\vee$" will be used for a generated line or plane; that is, $P \vee Q$ will give the line generated by the points $P$ and $Q$ while $R \vee a$ denotes the plane generated by the point $R$ and the line $a$. The symbols $\wedge$ and $\vee$ will also be used in logical statements; for example $A \wedge a = \emptyset$ means that the intersection of the point $A$ and the line $a$ is the empty set, that is the point $A$ does not belong to the line $a$. Both the relation "$\wedge$" and "$\vee$" are associative and, consequently, their meaning can be extended to more than two operands. This means for example that the notation $P \vee Q \vee S$ can be used to denote the plane generated by the points $P$,$Q$ and $S$. Finally, $\mathbb{E}^2$ will be used to denote the 2D Euclidean plane in general whereas $\mathbb{E}^3$ will be the Euclidean (three dimensional) space.

The set of real numbers will be denoted by $\mathbb{R}$ and the symbols $\mathbb{R}^2$, $\mathbb{R}^3$, $\mathbb{R}^n$ etc. will denote the set of *column* vectors of corresponding dimensions. The vectors themselves will be denoted by small Latin characters as well, choosing characters usually at the end of the alphabet. If $x \in \mathbb{R}^3$, $x^T$ denotes the *transpose* of the vector $x$. To save space, $(1,2,3)^T$ will be used instead of

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{2.1}$$

To resolve ambiguities, the notation $\vec{x}$ will also be used for a vector to distinguish it from a line. In a number of cases a coordinate system will be implicitly present in the geometric environment in use. In such cases, the points will be identified with their coordinate vector and characters like $p$ or $q$, which denote in fact vectors, will also be used to denote points. This convention, although not necessarily very precise mathematically, will be extensively used later.

Matrices will be denoted by capital Latin characters with their elements being the corresponding small characters (that is the elements of the matrix $A$ will be $a_{i,j}$). $A^T$ denotes the transposed matrix of $A$; if $x$ is a vector of the appropriate dimension, $Ax$ will be the (multiplied) column vector and $x^T A$ the row vector. In case of doubt, the notation $\bar{A}$ will also be used to differentiate matrices from points.

The scalar (or inner) product of two vectors $x$ and $y$ will be denoted by $x^T y$; the vector (or outer) product of two vectors is $x \times y$. This latter product can be calculated by evaluating the following formal determinant:

$$xxy = \det\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} \tag{2.2}$$

where $\mathbf{e}_i$, $(i = 1,2,3)$ denote the basic unit vectors of $I\!R^3$ (that is $(1,0,0)^T$, $(0,1,0)^T$ and $(0,0,1)^T$). In other words, the vector product is:

$$xxy = \left( \det\begin{bmatrix} x_2 & x_3 \\ y_2 & y_3 \end{bmatrix}, -\det\begin{bmatrix} x_1 & x_3 \\ y_1 & y_3 \end{bmatrix}, \det\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} \right)^T \tag{2.3}$$

If $A$ is a quadratic matrix of dimension $n$ and $x,y \in I\!R^n$, $xAy$ will denote the so called *bilinear form*, that is:

$$xAy = x^T(Ay) = (x^TA)y \tag{2.4}$$

## 2.3. The Axiomatic System of Projective Geometry

### 2.3.1. Background

Figure 2.1 illustrates what were the basic problems which led to the development of projective geometry. A central projection is made onto the plane $\Pi$; for the sake of simplicity we concentrate now on projecting (from the centre $C$) the plane $\Psi$ onto $\Pi$.

The central projection has a number of very nice properties as seen from the figure. It maps (almost) all points of $\Psi$ onto points of $\Pi$; it maps a line of $\Psi$ onto a line of $\Pi$ and maps (usually) the points of intersections of two lines onto the point of intersection of the image of these lines. It is also almost invertible; that is for almost all points of $\Pi$ there is a corresponding point of $\Psi$ which would be the inverse image.

Figure 2.1 also shows why such vague statements are to be used to characterise this mapping. Indeed, there are some points of $\Psi$ (eg $P$) for which the central projection is not properly defined (the projection line does not have an intersection with the image plane). Accordingly, all lines which intersect in points for which no image could be defined (like $m$ and $n$), though being mapped onto lines, become parallel on $\Pi$, that is they have no intersection points any more. Additionally, all points on the line $l'$ on $\Pi$ (which is the intersection line of $\Pi$ and a plane containing $C$ and parallel to $\Psi$) are without inverse image.

The reason for all these singularities can be traced back to the very existence of parallel lines in a Euclidean environment. One would like to interpret somehow what happens to the intersection point of parallel lines; if this were done, the image of $P$ could be defined as being the "intersection" in some sense of the parallel lines of $m'$ and $n'$. Clearly, the problem is that no Euclidean point can play such a role; there are "holes", or missing elements in the set of all points in a Euclidean plane.

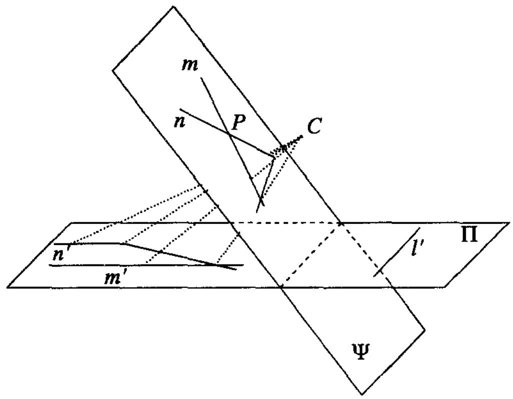



Figure 2.1.

What is usually done in such cases in mathematical practice is to *enlarge* the basic set one is working with. In other words, a new set is created which would contain (in this case) the set of all Euclidean planar points but which would also contain some additional elements. The usual notions (in this case "lines", "intersection" etc.) should be *extended* for this larger set so as to include the traditional notions as well. If this extension is well done, the "holes" may be filled and one arrives at a much clearer structure than the original one. This is what will be done for projective geometry: new elements (which are not Euclidean points) will be defined and the notion of lines, line intersections etc. will be extended so as to include these new elements as well, such that all the problems stated in figure 2.1 can be overcome. These new elements will be called the *ideal points*; they will be the mathematically precise form of what is commonly called "points at infinity".

The way of doing this extension is again very classic in mathematics but it requires some abstractions which are not always easy to understand for non–mathematicians. What has to be used is what is called an equivalence relation and the generated quotient set. This is as follows.

A (binary) *relation* on a set $H$ is defined to be a subset of the set of element pairs (that is a subset of $H \times H$). If $x,y \in H$ and the relation is denoted by $\rho$, than $x\rho y$ denotes the fact that the two elements $x$ and $y$ belong to the same subset, that is the defined relation "holds" for them. A relation $\rho$ is an *equivalence relation* if the following three properties hold:

$$\begin{aligned} &\forall x,y,z : \; x\rho y \wedge y\rho z \Rightarrow x\rho z \\ &\forall x,y \;\; : x\rho y \Rightarrow y\rho x \\ &\forall x \;\;\;\;\; : x\rho x \end{aligned} \tag{2.5}$$

The relation is said to be transitive, symmetric and reflexive. One can easily recognise that the relation "=" on $\mathbb{R}$ is an equivalence relation.

Equivalence relations divide the set on which they are defined into a set of mutually disjoint subsets (they provide an abstract tessellation of the supporting set). Indeed, if $A$ is an arbitrary set and $\rho$ is a relation defined on $A$ for which (2.5) holds, then for each $x \in A$ the following set can be defined:

$$x_\rho = \{ z : z \in A \text{ and } x \rho z \} \tag{2.6}$$

These sets are called the *equivalence classes* of $A$ defined by $\rho$. It is a typical mathematical exercise to prove that if $x \rho y$ holds then also $x_\rho = y_\rho$ and if $x \rho y$ does not hold then $x_\rho \cap y_\rho = \emptyset$. In other words, the equivalence classes are disjoint sets which are "generated" by each elements of the set $A$. As a result of the reflexivity, $x \in x_\rho$; that is, the equivalence classes effectively tessellate the whole set (no element is left out). Proving the previous statements is not particularly complicated; this is left to the reader.

As a result of the tessellation one can also speak of a new set, denoted by $A_\rho$, by taking:

$$A_\rho = \{ x_\rho : x \in A \} \tag{2.7}$$

This set is usually called the *quotient set* or *quotient space*. If one wants to go beyond the abstract definition, it could be said that it is a set of which the elements characterise the equivalence relation $\rho$ by collecting into one element all elements of $A$ which somehow belong together.

The quotient set is widely used for a mathematically precise formulation of an extension mechanism. An equivalence relation is defined either on the elements of the set $A$ or on some other set $B$ related strongly to $A$; the set $A \cup B_\rho$ provides then an extension of $A$ which makes the characterisation of the relation $\rho$ simpler (provided the set $B$ is chosen in an appropriate manner). This approach is very widespread in mathematics; this is how for example irrational numbers are constructed out of rational ones on $\mathbb{R}$, and this is also how ideal points are defined properly.

### 2.3.2. The Basic Construction

For the sake of simplicity, in what follows the axiomatic system for a projective plane will be described systematically; the construction leading to the projective space is quite similar and only the major differences will be presented. As shown later, there exist very good means to give an intuitive picture of a projective plane whereas it is much more difficult to visualise a projective space; consequently, projective planes will always be used as illustrative examples even if the real environments used in graphics systems tend to be a projective space rather than a plane.

If $\mathbb{E}^2$ is the Euclidean plane, let us denote by $\Lambda(\mathbb{E}^2)$ the set of all lines of $\mathbb{E}^2$. On this set, the relation of *parallelism* is an equivalence relation (provided that each line is considered to be parallel to itself). In other words, the relation $\rho$ could be defined by:

$$\forall n,m \in \Lambda(\mathbb{E}^2): n\rho m \leftrightarrow n \text{ and } m \text{ are parallel} \tag{2.8}$$

The fact that this relation is an equivalence relation can be seen easily. Consequently, one may speak of the quotient space of $\Lambda(\mathbb{E}^2)$, which (instead of $\Lambda(\mathbb{E}^2)_\rho$) will be simply denoted by $\mathbb{I}$.

Intuitively speaking the elements of $\mathbb{I}$ are mathematically precise abstractions of what is common in two lines of $\mathbb{E}^2$ vis–a–vis parallelism. Indeed, two parallel lines will generate the very same element of $\mathbb{I}$ (using formula (2.6)) and elements generated by two non–parallel lines will be different. This also means that if a new set is defined by

$$\mathbb{P}E^2 = \mathbb{E}^2 \cup \mathbb{I} \tag{2.9}$$

the resulting set will contain on the one hand the "traditional" Euclidean points plus some abstract elements describing the "common part" of two parallel lines of $\mathbb{E}^2$.

To define some kind of geometry on $\mathbb{P}E^2$, the notions of points, lines, intersections of lines etc. have to be extended onto this larger set; of course only an extension which would preserve the "traditional" Euclidean notions is of real interest. By defining these extensions and by finding some elementary properties of them, a new mathematical structure, a new *geometry* will be created. Mathematically, this means that the set of elementary properties might also be considered as a new set of axioms (much the same way as the axioms described in the Elements of Euclid form the basis of Euclidean geometry or the modified set of these axioms defined by Bólyai and Lobatchewsky would form the axiomatic basis for hyperbolic geometry). This new geometry is called *projective geometry*.

In Euclidean geometry, the notion of point is just another name for the (set–theoretical) elements of the set $\mathbb{E}^2$ (or $\mathbb{E}^3$). The same approach can be used in the case of projective geometry, that is:

> **Definition 2.1.** The elements of $\mathbb{P}E^2$ are called *(projective) points*. In case it is necessary to make a difference, the elements of $\mathbb{I}$ are also called *ideal points* whereas the elements of $\mathbb{E}^2$ are also called *affine points*†.

Lines in Euclidean geometry are special subsets of $\mathbb{E}^2$; the properties of these subsets are described in the axioms of the Euclidean axiomatic system (the notation $\Lambda(\mathbb{E}^2)$ has been used to denote the set of all lines). The aim is to maintain Euclidean lines in the new environment as well. Here is the precise definition:

†The term *directions* is also in use for affine points.

**Definition 2.2.** *Lines* of $\mathbb{P}E^2$ are special subsets of $\mathbb{P}E^2$. The set of all lines is denoted by $\Lambda(\mathbb{P}E^2)$ and its elements can be described as follows:

$$\Lambda(\mathbb{P}E^2) = \{x \cup \{x_\rho\} : x \in \Lambda(\mathbb{E}^2)\} \cup \{\mathbb{I}\}$$

where $\rho$ stands for the equivalence relation "parallelism".

In plain English: each line of $\mathbb{E}^2$ is extended by its direction, that is the ideal point generated by the line using (2.6); additionally, the set of all ideal points is also considered as a line. Just as in the case of points, if there is a necessity to make a difference, $\mathbb{I}$ will also be called the *ideal line* (there is only one such line!) whereas all other lines are the *affine lines*. An affine line is *not* equal to a Euclidean line; it is a Euclidean line *plus* one point (also called the ideal point of the line).

The intersection of a line and a point is just another terminology for the set-theoretical inclusion; it is therefore automatically valid for the new environment as well.

What are the basic properties of these lines and points (also called, to make the distinction, projective lines and points)? There are some statements forming a set of basic theorems on these notions and which are as follows.

**Theorem 2.1.** For each two elements of $\mathbb{P}E^2$ the following holds:

$$\forall P,Q \in \mathbb{P}E^2, (P \neq Q): \exists!\, l \in \Lambda(\mathbb{P}E^2) \text{ for which}$$
$$P \wedge l \neq \emptyset \text{ and } Q \wedge l \neq \emptyset$$

That is for every pair of projective points there exists one and only one projective line which contains the given two points.

The proof of this theorem follows a fairly standard mathematical line of thought: different cases should be examined apart.

(1) If both $P$ and $Q$ are affine points, the corresponding axiom of the Euclidean geometry says that there is one and only one Euclidean line which intersects both $P$ and $Q$; the corresponding affine line will do for the projective case. It is trivial to see that no other projective line will satisfy the requirements.

(2) If both $P$ and $Q$ are ideal points, the ideal line will contain both of them; furthermore, no affine line may intersect two distinct ideal points.

(3) Finally if $P$ is affine and $Q$ ideal, there is a whole set of affine lines intersecting $P$. However, out of these lines only one may have as an ideal point $Q$: the ideal point is just an element of $\mathbb{I}$ determined by the relation (2.6). ∎

An analogous statement for lines is as follows.

**Theorem 2.2.** For every pair of lines on $I\!PE^2$ the following holds:

$$\forall l, n \in \Lambda(I\!PE^2)\ (l \neq n) \colon \exists!\ P \in I\!PE^2 \text{ for which}$$

$$P \wedge l \neq \emptyset \text{ and } P \wedge n \neq \emptyset$$

That is each two distinct lines have an intersection point (there are no parallel lines!). This intersection point is denoted by $l \wedge n$. The proof of this theorem is similar to the previous one:

(1) If both $n$ and $m$ are affine, there are again two cases:

  (a) the two lines are parallel in the Euclidean sense; in this case they share the same ideal point (according to (2.6));

  (b) the two lines have an Euclidean intersection point which will also serve as an intersection point in the projective sense as well.

(2) If $n$ is affine and $m$ is the ideal line, the ideal point of $n$ (which exists according to definition 2.2) is the intersection point. ■

Finally, two theorems are needed which are rather technical in nature but are necessary for the full description of the whole theory:

**Theorem 2.3.** Each projective line contains at least three points.

**Theorem 2.4.** There exist three points on $I\!PE^2$ which are not on the same line (they are not *collinear*).

Theorems 2.1 to 2.4, together with the corresponding definitions 2.1 and 2.2 form the axiomatic foundation of (planar) projective geometry. By using these notions and theorems the ambiguities of the description related to figure 2.1 may now be removed. The planes $\Pi$ and $\Psi$ should now be considered projective planes instead of Euclidean ones. $P$, which is an affine point of $\Psi$, will be mapped onto an ideal point of $\Pi$; the lines $n$ and $m$ which intersect at $P$ on $\Psi$ will be mapped onto $\Pi$ (that is onto $n'$ and $m'$) by still keeping the line intersection; the only problem is that this intersection point happens to be an ideal point. In the case of projective geometry, this makes no real difference, however.

A *projective space* can be constructed very similarly. Instead of $\Lambda(I\!E^2)$, $\Lambda(I\!E^3)$ should be considered for the equivalence relation; the resulting quotient set will contain the set of *ideal points* again. The set of all projective points will be

$$I\!PE^3 = I\!E^3 \cup I\!I \tag{2.10}$$

just as in the case of the plane.

The definitions corresponding to 2.1 and 2.2 are very similar but there are however some differences. In the case of spatial geometry, there are two special kinds of subspaces: lines and planes. Likewise, projective lines and projective planes should both be defined to ensure a correct extension.

The most important remark which helps to make these extensions possible is

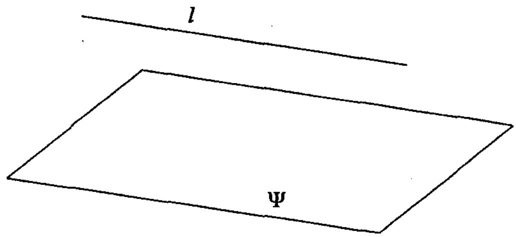



Figure 2.2.

as follows. If $\Pi(\mathbb{E}^3)$ denotes the set of all planes of $\mathbb{E}^3$ and $\Psi\in\Pi(\mathbb{E}^3)$, this plane generates a special subset of $\mathbb{I}$ (see also figure 2.2). Indeed, the ideal points which are parallel to $\Psi$ (like the one generated by $l$ on figure 2.2) will have a special characteristic: they will describe all the planes and lines which are parallel to $\Psi$. Denoting this subset of $\mathbb{I}$ by $\Psi_p$, the following definition may be accepted:

**Definition 2.3.** *Lines* of $\mathbb{P}E^3$ are special subsets of $\mathbb{P}E^3$. The set of all lines is denoted by $\Lambda(\mathbb{P}E^3)$ and its elements may be described as follows:

$$\Lambda(\mathbb{P}E^3) = \{x\cup\{x_p\} : x\in\Lambda(\mathbb{E}^3)\} \cup \{\Psi_p : \Psi\in\Pi(\mathbb{E}^3)\}$$

In other words, each spatial plane induces a subset of $\mathbb{I}$ which is called an *ideal line*; there is one such ideal line for each set of parallel planes. It is intuitively clear (and mathematically easily provable) that the equivalence classes induced on $\Pi(\mathbb{E}^3)$ by the relation parallelism are isomorphic to these ideal lines.

**Definition 2.4.** Planes of $\mathbb{P}E^3$ are special subsets of $\mathbb{P}E^3$. The set of all planes is denoted by $\Pi(\mathbb{P}E^3)$ and its elements may be described as follows:

$$\Pi(\mathbb{P}E^3) = \{\Psi\cup\{\Psi_p\} : \Psi\in\Pi(\mathbb{E}^3)\} \cup \{\mathbb{I}\}$$

Here again, $\mathbb{I}$ is called the *ideal plane* whereas all other planes are denoted by *affine planes*. Definitions 2.1, 2.3 and 2.4 form the necessary extensions for projective space. There is a set of theorems which form the necessary axiomatic basis for the theory. The theorems themselves are very much the same as above in flavour and they are just listed here without proof (the reader may easily reproduce the mathematical deductions). These theorems are as follows (some very technical ones like the analogies of theorem 2.4 are omitted).

**Theorem 2.5.** If a point is the element of a line and the line is a subset of a plane then the point is an element of the plane (that is the two different kind of subsets in $\mathbb{P}E^3$ form a "hierarchy").

**Theorem 2.6.** Every pair of points generate one and only one line which contains them both (denoted by $P \vee Q$).

**Theorem 2.7.** Every pair of planes have one and only one common intersection line (denoted by $\Pi \wedge \Psi$). In other words, there are no parallel planes.

**Theorem 2.8.** A line and a plane has either an intersection point or the line belongs to the plane.

**Theorem 2.9.** For every pair of lines there is at most one plane which contains them both (denoted by $l \vee n$).

**Theorem 2.10.** Each three non collinear points generate one and only one plane which contains them all (denoted by $P \vee Q \vee S$).

**Theorem 2.11.** Each three planes intersect in either a line or one point (denoted by $\Pi \wedge \Psi \wedge \Phi$).

**Theorem 2.12.** If two lines are coplanar (that is there exists a plane which contains them both), they have one and only one intersection point (denoted by $l \wedge m$).

One has to be very careful in the case of the last statement: in projective space there may be lines which do not have an intersection point (these kinds of lines are not considered to be parallel in classical Euclidean geometry either). It is true, however, that there are no parallel lines any more. Another point of interest: each plane in $\mathbb{P}\mathbb{E}^3$ can be considered as a projective plane by itself (just as each plane in $\mathbb{E}^3$ behaves "locally" as a Euclidean plane). This statement sounds trivial but in a precise formulation of projective geometry it has to be proven that the axioms of the projective plane are valid locally as well.

As said before, theorems 2.1 up to 2.4 together with the definitions 2.1 and 2.2 (or their three–dimensional counterparts), may be considered as an axiomatic system: this is the axiomatic foundation of projective geometry. In theory, based on these axioms, all the (sometimes obscure) steps of the construction may be relegated to the background: one could just speak of points and lines in $\mathbb{P}\mathbb{E}^2$ where no parallel lines exist any more. This geometry has a very different nature compared to the "well–known" Euclidean one. A projective plane or space is locally very much like its Euclidean counterpart but has different global characteristics (the exact relationships between Euclidean and projective geometry will become clear in a later chapter).

The main difference comes from the fact that a projective line behaves much like an Euclidean (planar) circle; in fact, it is isomorphic with it. The ideal point of the line is the element which somehow "glues" the two ends. This fact has far reaching consequences: the very notion of line segment has no meaning any more (by giving the points $A$ and $B$ on the line, one can reach $B$ from $A$ in two ways). Consequently, the concept of the interior of a polygon disappears from the theory as do convex polygons.† There is no way of defining the notions of "clockwise" and

†More generally, Jordan's theorem, which states the existence of the interior and the exterior of a planar area generated by a "well-behaved" curve on the Euclidean plane is not true any more.

"anticlockwise" on a projective plane; mathematically speaking, the projective plane is not orientable.

Of course, these differences are at the source of a number of problems when projective geometry has to be used for the purposes of computer graphics. Computer graphics applications rely heavily on, for example, the interior of a polygon, which also forms an integral part of all ISO standards on graphics (see all ISO documents in the references). One possibility would be to avoid the use of this theory; however, as stated in the introduction, this is barely possible. Another approach would be to develop an alternative mathematical theory which would try to avoid the appearance of these problems; an attempt has been made recently by J. Stolfi ([Stol89]) based on some earlier mathematical works made by H. Grassmann about a hundred years ago and followed by a number of other mathematicians (Stolfi refers to the works [Berm61] and [Hest84] for earlier references). In his thesis, Stolfi describes the theory of *oriented projective spaces*, which contains many similarities to classical projective geometry but where the notion of the orientation of a line, plane and space still has a meaning. However, the mathematical theory and formulae involved tend to be rather complicated and quite abstract; to use it in practice would probably require a reformulation of a number of classical approaches which have been in use in computer graphics in the past 10 to 15 years. Whether this is worthwhile or not is still to be proven; the approach is of interest, however.

The approach described in this thesis is much more pragmatic. Projective geometry should be used, because it is a precise description of practical problems arising in computer graphics and it also helps to create more efficient algorithms and methods. However, the extreme axiomatic nature of projective geometry, which would ignore the origins of, for example, ideal points will not be followed everywhere; in most of the cases the construction described here will be present in the background. By carefully exploiting the relationships between projective geometry and the Euclidean one, some of the problems may be described in terms of a Euclidean environment even if the price to be paid might be sometimes to use four dimensional geometry instead of the well known two and/or three dimensional ones. In some other cases the full power of projective geometry has to be used (eg for the handling of conics). By alternating between projective and Euclidean geometries, most of the problems can be avoided in a down–to–earth but still powerful way.

In contrast to the traditional and "purist" projective geometry textbooks, the numerical aspects of projective geometry are of primary interest to computer graphics scientists. Some kind of coordinate system is essential to be able to describe geometric entities with numbers and hence make them manageable by computers. This is will be covered in the next section.

## 2.4. Projective Coordinate Systems (Homogeneous Coordinates)

Coordinate systems as used in Euclidean geometry were only introduced in the $18^{th}$ century. Their use has become so natural that one tends to underestimate the importance and the mathematical difficulties involved when using them. The use of the

Cartesian system creates a "bridge" between two very different mathematical theories, namely Euclidean geometry and the theory of real numbers. A more exact mathematical formulation of what the Cartesian coordinate system really means is presented here, to show what is the necessary approach to achieve something analogous for projective plane/space.

**Theorem 2.13.** If $O$, $E_1$ and $E_2$ are three non-collinear points of the Euclidean plane $\mathbb{E}^2$, then there exists a one-to-one correspondence between $\mathbb{E}^2$ and $\mathbb{R}^2$ so that the point $O$ will correspond to the vector $(0,0)^T$, the point $E_1$ to $(1,0)^T$ and, finally, $E_2$ to $(0,1)^T$. If, furthermore, it is required that the distance of the points $P$ and $Q$ should be expressed by the formula:

$$dist(P,Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$$

(where $P$ is mapped onto $(p_1,p_2)^T$ and $Q$ onto $(q_1,q_2)^T$), then there exists only one such correspondence which fulfils these requirements.

It is not possible to have a one-to-one correspondence between the projective plane and $\mathbb{R}^2$. A mapping is however provided by the use of *homogeneous coordinates*.

Two non-zero vectors $a,b \in \mathbb{R}^n$ are considered to be equal in the homogeneous sense if there exists a non-zero $\lambda \in \mathbb{R}$ so that the equality $a = \lambda b$ holds. This relation is an equivalence relation on $\mathbb{R}^n - \{0\}$ (the origin). The corresponding quotient set (denoted by $\mathbb{PR}^n$ in the following discussion) is called the set of homogeneous vectors. The same vector notations will be used to denote its elements, but their homogeneous nature must always be kept in mind. In case of doubt, the notation $[(a_1,a_2,\ldots,a_n)]^T$ will also be used to denote the homogeneous vector generated by $(a_1,a_2,\ldots,a_n)^T \in \mathbb{R}^n$.

The theorem which is analogous to 2.13 is as follows.

**Theorem 2.14.** If $O$, $A_1$, $A_2$ and $E$ are elements of $\mathbb{PE}^2$, so that no three of them would be collinear (they are of a *general position*), then there exists a unique one-to-one correspondence between the points of $\mathbb{PE}^2$ and the elements of $\mathbb{PR}^3$ so that the following relations

$$\begin{aligned} O &\leftrightarrow [(0,0,1)]^T \\ A_1 &\leftrightarrow [(1,0,0)]^T \\ A_2 &\leftrightarrow [(0,1,0)]^T \\ E &\leftrightarrow [(1,1,1)]^T \end{aligned} \tag{2.11}$$

hold.

For a projective space, one more point ($A_3$) is necessary; the requirement is not only that there should be no three collinear points but also that there should not be four coplanar points. The mapping is then performed between $\mathbb{P}E^3$ and $\mathbb{P}R^4$ and (2.11) becomes:

$$\begin{aligned} O &\leftrightarrow [(0,0,0,1)]^T \\ A_1 &\leftrightarrow [(1,0,0,0)]^T \\ A_2 &\leftrightarrow [(0,1,0,0)]^T \\ A_3 &\leftrightarrow [(0,0,1,0)]^T \\ E &\leftrightarrow [(1,1,1,1)]^T \end{aligned} \tag{2.12}$$

Traditionally, the coordinate components of the homogeneous vectors are denoted either by subscripted Latin characters or by using the letter $w$ for the last coordinate eg of the form $[(x,y,w)]$. The exact proofs of the theorems 2.13 and 2.14 would go far beyond the scope of the present thesis; the interested reader should consult, for example, [Keré66].

The exact relationships between the Cartesian coordinates and the homogeneous ones play an essential role in the following sections. Indeed, the geometric environment which is the usual starting point for any graphics system is a Euclidean one together with some coordinate system defined on $\mathbb{E}^2$ or $\mathbb{E}^3$; a coordinate system which would be effective in some sense for the object which is to be described. This Euclidean environment has to be treated, however, as a projective one by the graphics system; there is therefore an extension to be made (described in the previous section) which would embed this plane or space into a projective plane or space respectively. This process is performed by adding ideal points to $\mathbb{E}^2$ or $\mathbb{E}^3$ to result in $\mathbb{P}E^2$ or $\mathbb{P}E^3$ respectively. The question is, which homogeneous coordinate system to use so that the relationship between the Cartesian and the homogeneous coordinates of a Euclidean (that is affine) point would be as simple as possible? This is done as follows.

**Theorem 2.15.** Suppose that a Cartesian coordinate system has been chosen on $\mathbb{E}^2$ (for the sake of simplicity the planar situation will be examined in detail first). Let the points $O, A_1, A_2$ and $E$ be defined as follows.

- Let $O$ be the origin of the Cartesian system;
- Let $A_1$ be the ideal point of the $x$ axis;
- Let $A_2$ be the ideal point of the $y$ axis;
- Let $E$ be the affine point with coordinates $(1,1)^T$

then, if the Cartesian coordinates of a point $P$ are denoted by $(p_1, p_2)^T$ the following relation holds:

If $P \in \mathbb{E}^2$, the homogeneous coordinates of $P$ are given by $[(p_1, p_2, 1)]^T$.

Furthermore, if $Q$ is an ideal point of $\mathbb{P}E^2$, it can be described by the line $Q \vee O$. With the help of such a line the following relationship also holds:

If $R \in Q \vee O$, the homogeneous coordinates of $Q$ are given by $[(r_1, r_2, 0)]^T$.

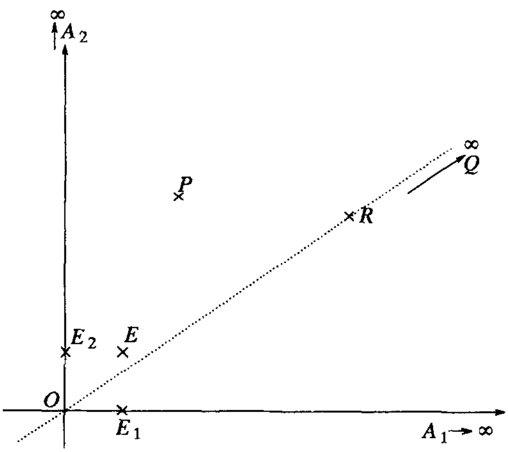



Figure 2.3.

It is fairly straightforward to see that the relations listed in theorem 2.15 do define a one–to–one mapping of $\mathbb{P}E^2$ and the set of homogeneous vectors. Taking into account the uniqueness statement formulated in theorem 2.14, the validity of theorem 2.15 follows easily. ■

Theorem 2.15 also means that *in this coordinate system* ideal points can be uniquely characterised by having the last coordinate value being zero. This also leads to the following formula (well known in computer graphics, see also figure 2.3):

If $P \in \mathbb{P}E^2$, $P$ is affine and a homogeneous coordinate system has been chosen for $\mathbb{P}E^2$ as described above, the formula

$$[(p_1, p_2, p_3)]^T \rightarrow (p_1/p_3, p_2/p_3)^T \tag{2.13}$$

will give the Cartesian coordinate values of $P$ (the fact that $P$ is affine is equivalent to the fact that $p_3 \neq 0$). In the computer graphics literature this step is usually called the "projective division"$^\dagger$.

It has to be stressed that such a unique characterisation of an ideal point by its

$^\dagger$The terms "perspective division" or "w-divide" are also in use.

coordinate values is possible only in this coordinate system. It is perfectly possible to choose another homogeneous coordinate system where this characterisation is not valid any more‡. Just as in the case of Cartesian coordinate systems, it is a fairly widespread approach, when trying to prove some theorem in projective geometry, to choose a coordinate system which fits the original problem itself (this approach is particularly fruitful when dealing with conics).

Homogeneous coordinates also provide a way of characterising lines (on $\mathbb{P}E^2$) and planes (on $\mathbb{P}E^3$). In the case of Cartesian coordinates, a line can be expressed by the equation:

$$ax_1 + bx_2 + c = 0 \tag{2.14}$$

with appropriate constants $a,b$ and $c$. By using the identification procedure for Cartesian versus homogeneous coordinates, this equation could be rewritten for homogeneous coordinates as follows:

$$ax_1 + bx_2 + cx_3 = 0 \tag{2.15}$$

It is also clear that if the values $a,b$ and $c$ are multiplied by any non-zero real number, equation (2.15) would still remain valid. This means that $[(a,b,c)]^T \in \mathbb{P}R^3$ gives an adequate description of a line. In other words, by using homogeneous coordinates, not only points but also lines can be assigned a homogeneous vector, which could just be called the homogeneous coordinates of a line (and the points of the line can be described by (2.15)). In case of $\mathbb{P}E^3$ planes can be described similarly (but not lines).

It can be proven that the above characterisation of lines/planes does not depend on the special choice of the homogeneous coordinate system used to derive (2.15). The fact that lines/planes can be described by homogeneous coordinates just like points leads to an elegant symmetry of all the formulae involving intersection points, generated lines etc.; they will be presented later. This similarity, which is referred to as the *duality principle* of projective geometry has, in fact, a much deeper background. By looking at the axioms of $\mathbb{P}E^2$ (or respectively of $\mathbb{P}E^3$), there is a similarity of the behaviour of points and lines: if the words/terms "point" are exchanged with "line", and "intersection points of lines" with "lines generated by points", valid statements will result. Consequently, this fact is also true not only for the axioms but for all statements derived from them. By making use of this duality principle a number of formulae can be derived easily and one might also get a clue for finding additional and useful formulae (see for example [Arok89]).

The use of homogeneous coordinates has been an accepted practice in computer graphics for a very long time; their description can be found in all "classical"

‡In fact, the whole approach might also be turned upside down. Indeed, if an arbitrary homogeneous coordinate system is given on $\mathbb{P}E^2$ or $\mathbb{P}E^3$, one could *define* the ideal points *of this coordinate system* to be the points with the last coordinate value being zero and all other ones being affine.

textbooks ([Newm79], [Fole84], [Salm87], [Fole90] and others) and more systematic descriptions can also be found in [Reis81] or [Bez83]. In most of these cases however, homogeneous coordinates are presented as being some kind of neat (one could even say "tricky") way of describing points so as to have a unified description of the effect of different transformations. While the usability of homogeneous coordinates even in 2D graphics is undeniable, it is important to realise that their use can be traced back to much more fundamental mathematical properties of projective geometry, which, in turn, plays a basic role in 3D graphics.

## 2.5. Isomorphic Models of Projective Planes and Projective Spaces

The use of homogeneous coordinates provides a means of "visualising" a projective plane (and to a smaller extent a projective space). The idea is to give some kind of an intuitively manageable surface in Euclidean geometry which would, in some sense, be a good model for a projective environment.

A homogeneous coordinate (that is an equivalence class) can be viewed as a line in a higher dimensional Euclidean space. That is, an element of $\mathbb{P}\mathbb{R}^3$ can be identified with a line in $\mathbb{R}^3$ crossing the origin†. This fact is clear from the definition of a homogeneous coordinate.

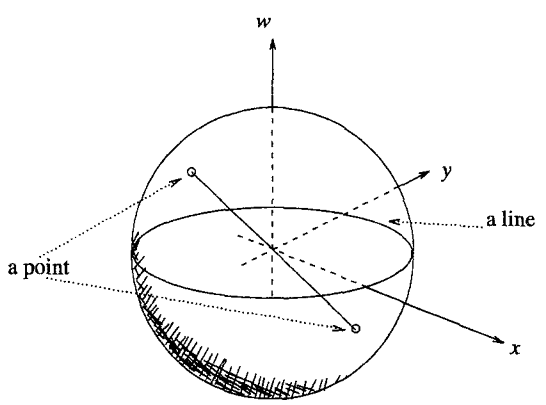



Figure 2.4.

### 2.5.1. The Riemann Sphere

The Riemann sphere (figure 2.4; also called the spherical model of a projective plane) is the unit sphere around the origin of $\mathbb{R}^3$ where all diametrically opposite points are identified. Each point in $\mathbb{P}\mathbb{E}^2$ is represented by a poins of the Riemann sphere, and lines of $\mathbb{P}\mathbb{E}^2$ are represented by the great circles on the sphere (again,

†More precisely, the origin should be removed from the line to get the exact identification.

with opposite points identified). The ideal line is represented by the great circle defined by $w=0$. With this mapping of $\mathbb{P}E^2$ onto the Riemann sphere, the latter becomes isomorphic to $\mathbb{P}E^2$. The model also shows the remarkable identity of an affine line and the ideal line.

The Riemann sphere is, although intuitively very helpful, not really of use in the forthcoming. The reason is that the identification of a sphere point with the corresponding affine point of $\mathbb{P}E^2$, and also with its Cartesian coordinates, leads to disagreeable formulae. It is, however, a very helpful tool to get new ideas; indeed, as opposed to the so called "straight model" of the projective plane (presented in the next section) the full plane, including the ideal line, is modelled by it.

The Riemann sphere can be generalised for projective space as well; one should take a unit sphere in $\mathbb{R}^4$ around the origin However, because of the difficulties of visualising a four dimensional space, this version of the Riemann sphere is not really useful.

### 2.5.2. Embedding into $\mathbb{R}^3/\mathbb{R}^4$ (the Straight Model)

The second, and more widespread model of a projective plane (the so-called straight model) is shown in figure 2.5. By using the identification of Cartesian and homogeneous coordinates, all *affine* points (that is the points of $\mathbb{E}^2$) may be represented by points of $\Pi$ (that is the plane $w=1$). This is clearly nothing else than a pictorial representation of the fact that homogeneous coordinates describe lines in a higher dimensional space. From a purely mathematical point of view, the drawback of this model is the fact that only the affine points can be represented so clearly. On the other hand, as far as computer graphics is concerned, the real issue is always to see how affine points are transformed; ideal points are just disagreeable but necessary additions to them. In this sense, the fact that the straight model shows only the affine points so clearly may well be an advantage rather than a disadvantage.

On the straight model ideal points are represented by homogeneous coordinates with the last coordinate value being zero. This means that ideal points are represented by lines running in the plane $x-y$, crossing the origin and having therefore no intersection points with $\Pi$ (see figure 2.5 again).

The fact of having chosen $\Pi$ to represent the projective points was, although a direct representation of the Cartesian-homogeneous identification, intentional: as said above, the fact that ideal points are so well separated from the affine ones might be helpful. Clearly, any plane could have been chosen equally well, like the plane $\Psi$ in figure 2.6. In this case the ideal points do become Euclidean points on $\Psi$ but, on the other hand, some of the affine points cannot be represented properly.

Like the Riemann sphere, this model works analogously for $\mathbb{P}E^3$: one should take the $w=1$ hyperspace in $\mathbb{R}^4$ to model $\mathbb{P}E^3$. Of course, the same problem arises: it is not possible to visualise properly the four dimensional space. This is the reason why figures 2.5 or 2.6 will be used in all cases when a pictorial representation will be necessary even if the real problems to be solved will be in $\mathbb{P}E^3$ rather than in $\mathbb{P}E^2$.

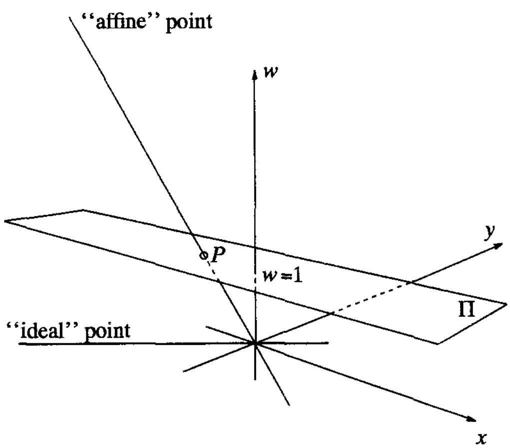



Figure 2.5.

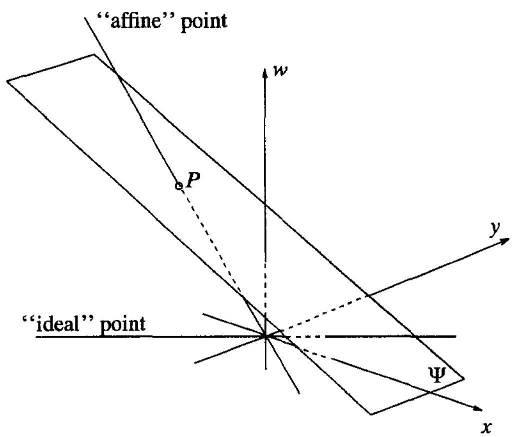



Figure 2.6.

There is one common point which has to be stressed in all forms of the straight model. In all these cases the projective environment has been embedded into a Euclidean environment again (with some restrictions). The significant difference is, however, that a Euclidean space of higher dimension was necessary. This will have an importance in what follows: by having first embedded the original Euclidean environment into a projective one and, in a second step, having re–embedded it into a Euclidean space of a higher dimension, many of the problems can be restated in a ''classical'' Euclidean way. This fact has never been really exploited in classical projective geometry; indeed, the problems arising for mathematicians are of a different nature. However, for computer graphics, this approach (exploited first in [Herm87]) has led to significant simplifications of a number of problems. Examples will be seen in later chapters.

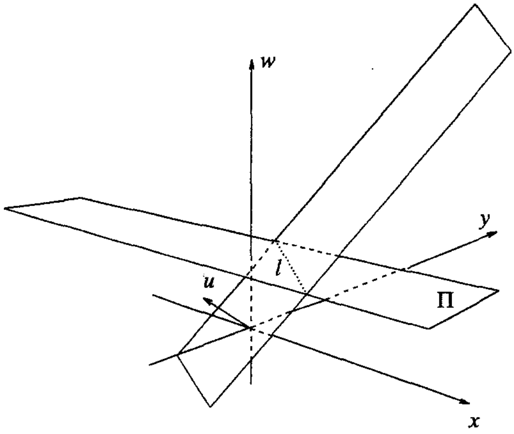



Figure 2.7.

The straight model also provides a way to represent all affine lines or planes (see figure 2.7). Let $u \in \mathbb{P}\mathbb{R}^3$ (or, for $\mathbb{P}\mathbb{E}^3$, $u \in \mathbb{P}\mathbb{R}^4$) be the homogeneous representation (that is the coordinates) of the line $l$. As an Euclidean vector, $u$ will also determine a plane in $\mathbb{R}^3$ (resp. a hyperspace in $\mathbb{R}^4$), namely the one containing the origin and whose normal vector is $u$. This plane (hyperspace) will intersect $\Pi$ in a line (resp. a plane) if $l$ is affine; the formula describing a line in homogeneous coordinates (that is (2.15)) simply proves that this intersection line/plane will just be $l$ itself. In other words, each affine line can be viewed as a plane (or a hyperspace) crossing the origin in $\mathbb{R}^3$ (resp. $\mathbb{R}^4$) if the straight model is used.

## 2.6. Some Basic Calculation Formulae

In what follows, a number of formulae (for the intersection of lines etc.) will be presented; these formulae will be used later. This is by no means an exhaustive list of all possible calculation methods offered by homogeneous coordinates; the reader should refer to the projective geometry textbooks and primarily to [Penn86] for further examples.

### 2.6.1. Formulae for the Projective Plane

If $u \in I\!PR^3$ and $v \in I\!PR^3$ represent two lines, the homogeneous coordinates of $u \wedge v$ are given by the following (formal) determinant:

$$u \wedge v = \det \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} \tag{2.16}$$

where $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ denote the vectors $(1,0,0)^T$, $(0,1,0)^T$ and $(0,0,1)^T$ respectively.

If $p \in I\!PR^3$ and $q \in I\!PR^3$ represent two points, the homogeneous coordinates of $p \vee q$ are given by the following (formal) determinant:

$$p \vee q = \det \begin{bmatrix} p_1 & p_2 & p_3 \\ q_1 & q_2 & q_3 \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} \tag{2.17}$$

To describe $p \vee q$, a parametric equation is sometimes more suitable than the description with homogeneous coordinates. The following formula is also true:

$$p \vee q = \{ \lambda p + \mu q \} \tag{2.18}$$
$$\lambda \neq 0 \text{ or } \mu \neq 0$$

The previous formulae may be used for slightly more complex calculations; for example to find the intersection point of two lines, knowing two points on each of them (a repetitive application of (2.17) and then (2.16) will do).

It is also important to know that in the case of $I\!PE^2$ the coordinates of some special points are known "by default". As examples, at least two distinct ideal points are known by their coordinates ( eg $[(1,0,0)]^T$ and $[(0,1,0)]^T$), the homogeneous coordinate of the ideal line is also known (it could be calculated by using (2.17) and the two distinct ideal points but it is also clear that $[(0,0,1)]^T$ should be the result).

It can be of interest to see what the ideal point of a given line is, provided that two of its affine points, say $p \in I\!PR^3$ and $q \in I\!PR^3$, are given. This can easily be calculated if the "straight model" is made use of: indeed, the direction of a line parallel to $\Pi$ and the (Euclidean!) line $p' \vee q'$ is to be found. However, if $p$ and $q$ are considered to be Euclidean, the vector from $q'$ to $p'$ can be calculated by:

$$\begin{bmatrix} p_1/p_3 \\ p_2/p_3 \\ 1 \end{bmatrix} - \begin{bmatrix} q_1/q_3 \\ q_2/q_3 \\ 1 \end{bmatrix} \tag{2.19}$$

Clearly, the homogeneous coordinates generated by (2.19) will give the ideal point of $p \vee q$ (see also figure 2.8).

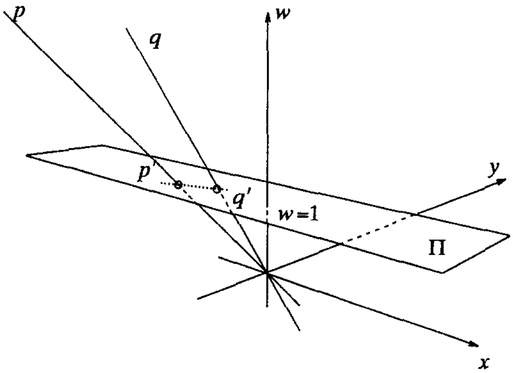



Figure 2.8.

### 2.6.2. Formulae for the Projective Space

If $u \in \mathbb{P}\mathbb{R}^4$, $v \in \mathbb{P}\mathbb{R}^4$ and $w \in \mathbb{P}\mathbb{R}^4$ represent three planes, the homogeneous coordinates of $u \wedge v \wedge w$ are given by the following (formal) determinant:

$$u \wedge v \wedge w = \det \begin{pmatrix} u_1 & u_2 & u_3 & u_4 \\ v_1 & v_2 & v_3 & v_4 \\ w_1 & w_2 & w_3 & w_4 \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 \end{pmatrix} \tag{2.20}$$

where $\mathbf{e}_1$, $\mathbf{e}_2$ $\mathbf{e}_3$ and $\mathbf{e}_4$ denote the vectors $(1,0,0,0)^T$, $(0,1,0,0^T)$, $(0,0,1,0)^T$ and $(0,0,0,1)^T$ respectively. If all subdeterminants in (2.20) are zero, the planes meet in a line; this is a singular case.

If $p \in \mathbb{P}\mathbb{R}^4$, $q \in \mathbb{P}\mathbb{R}^4$ and $r \in \mathbb{P}\mathbb{R}^4$ represent three points, the homogeneous coordinates of $p \vee q \vee r$ are given by the following (formal) determinant:

$$p \vee q \vee q = \det \begin{pmatrix} p_1 & p_2 & p_3 & p_4 \\ q_1 & q_2 & q_3 & q_4 \\ r_1 & r_2 & r_3 & r_4 \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 \end{pmatrix} \tag{2.21}$$

It is also worth mentioning that formula (2.18) remains valid in $\mathbb{P}\mathbb{E}^3$ as well, although the fundamental symmetry among formulae is now between points and planes instead of points and lines.

In $\mathbb{P}\mathbb{E}^3$ at least three distinct and non-collinear ideal points are known by their coordinates (eg $[(1,0,0,0)]^T$, $[(0,1,0,0)]^T$, $[(0,0,1,0)]^T$). The homogeneous vector $[(0,0,0,1)]^T$ gives the coordinates of the ideal plane.

If three non-collinear points of a plane Π are known, the homogeneous representation of two ideal points of $\Pi$ can be calculated analogously to (2.19). By using (2.18), the parametric representation of the ideal line on $\Pi$ can also be described.

## 2.7. Collinearities

In his already cited work, the *Erlanger Programm*, F. Klein has proposed the classification of different branches of geometry based on a class of transformations which would leave some part of the geometric structure invariant. Although the usual geometry textbooks do not follow this rigorous classification, to describe the properties of a geometry the description of an appropriate class of transformation might be very important.

In the case of projective geometry the geometrical structure is fully determined by lines, the intersection of lines and of the lines generated by points. It is therefore natural to accept the following definition to describe a basic set of transformations:

> **Definition 2.5.** A transformation $T: \mathbb{P}E^2 \rightarrow \mathbb{P}E^2$ (or $T: \mathbb{P}E^3 \rightarrow \mathbb{P}E^3$) is said to be a collinearity if for each three collinear points $P,Q,R \in \mathbb{P}E^2$ (or $P,Q,R \in \mathbb{P}E^3$ respectively) the points $T(P),T(Q),T(R)$ are also collinear.

Collinearities (also called *projective transformations* or *projective mappings*) play a significant role in projective geometry. These transformations map lines onto lines (according to definition 2.5), they map the intersection point of two lines onto the intersection point again and, finally, they also map the lines generated by points onto the line generated by the images of the points. In case of $\mathbb{P}E^3$, planes are also mapped onto planes as well as the intersection of planes onto the intersection of planes.

The theoretical importance of collinearities can be well understood from the fact that they keep invariant all properties used and described by the axioms of projective geometry. Indeed, all axioms are specified with the help of points, lines and intersections, properties which are invariant to projective mappings. The practical importance of these mappings is also high: "familiar" transformations like rotations, translations or scalings as well as central projections are all collinearities. Care should be taken, however, with the last example. Central projections are collinearities in the *projective sense* whereas they are not necessarily ones in the Euclidean sense; this was exactly the problem which had led to the use of this theory. Those readers who are familiar with the three dimensional graphics standards like GKS–3D, PHIGS ([ISO88], [ISO89], [ISO89a]), or related packages like PEX ([ISO88b], [Clif88]) can also realise that the notion of projective transformations encapsulates all possible transformations described in these documents, like modelling transformations and viewing. These examples will be detailed in what follows.

For practical and also theoretical purposes a very important sub-class of collinearities is the class of *affine transformations*. Its definition is as follows.

**Definition 2.6.** A collinearity is said to be an *affine* transformation if the images of all affine points remain affine. For non–singular transformations this also means that the images of all ideal points remain ideal.

Clearly, affine transformations are "closer" to Euclidean geometry than projective transformations in general; they leave the "dual" structure of a projective plane/space (ie the division among ideal and affine points) essentially intact. Among the examples cited above, rotations, translations and scalings are clearly affine while central projections are not.

A line (for $\mathbb{P}E^2$) or a plane (for $\mathbb{P}E^3$) is called the *vanishing line/plane* of the transformation if it is transformed onto the ideal line/plane respectively. A transformation is affine if and only if its vanishing line/plane is the ideal one.

One very important issue in projective geometry is to find the *projective invariant* features of geometric primitives or constructions. Projective invariance means that the given construction and/or the geometric features related to a given primitive would remain unchanged if a projective transformation were applied (more specific examples will be given later). Likewise, one could speak about *affine invariance,* related to features invariant for affine transformations.

Projective and affine invariance are not only of theoretical interest. In the case of computer graphics algorithms, one of the clues for a simplification or an improvement might be to find the projective/affine invariant part of them. As an example (and there will be much more later) one might think of the fact that a number of graphics primitives may be described in a very compact form with the help of some points only (eg conics), but for the final rendering some kind of a linear approximation of the primitive is necessary. It is of great importance to find projective/affine invariant representations of these primitives; such representations allow the postponement of the linear approximation along the graphics output pipeline, resulting therefore in faster rendering and better approximations.

The existence and uniqueness of collinearities is provided by the following theorem.

**Theorem 2.16.** If $\Pi$ and $\Psi$ are two projective planes (which may be identical), $P_i \in \Pi$ and $P_i' \in \Psi$ $(i=1,...,4)$ are points in $\Pi$ and $\Psi$ respectively such that no three of them are collinear, then there exists one and only one collinearity $T: \Pi \to \Psi$ for which $T(P_i) = P_i'$. For projective spaces five points are needed instead of four with the additional requirement that no four points may be coplanar.

Here again, the proof of this theorem would go beyond the scope of this thesis; the interested reader should consult for example [Keré66], [Coxe49] or [Penn86].

This theorem seems to have a theoretical value only, as far as computer graphics are concerned, but this is not absolutely true. It happens quite often that several, at first glance very different, methods are created to calculate an effective projective transformation. Theorem 2.16 provides a way to check whether the different approaches do generate the same mapping or not: only the images of four/five points have to be checked and the theorem ensures the uniqueness of the projective mapping provided that the images of these four/five points coincide.

### 2.7.1. Representation of Collinearities

As with projective points it is obviously of paramount importance to find some kind of numerical representation of collinearities. The corresponding theorem (which is one of the most important theorems in projective geometry) is as follows.

**Theorem 2.17.** If $\Pi$ and $\Psi$ are two projective planes (which may be identical) with a homogeneous coordinate system chosen on them and $T: \Pi \rightarrow \Psi$ is a collinearity, then there exists a 3x3 matrix $\bar{T}$ which describes the transformation as follows. If $x \in \mathbb{P}\mathbb{R}^3$ represents the point $X \in \Pi$ then

$$[\bar{T}x] \in \mathbb{P}\mathbb{R}^3 \tag{2.22}$$

gives the homogeneous coordinates of $T(X)$. Furthermore, $\bar{T}$ is uniquely defined in a homogeneous sense; that is, if $\bar{T}$ and $\bar{T}'$ both fulfil (2.22), then there exists a non-zero real number $\lambda$ for which $\bar{T} = \lambda \bar{T}'$.

In other words, the transformation can be described by a matrix-vector multiplication. For $\mathbb{P}\mathbb{E}^3$, the same theorem applies, with the obvious difference that the matrices involved are 4x4 rather than 3x3. In both cases, the singularity of the transformation is equivalent to the singularity of the corresponding matrix.

The opposite statement is also true, namely that formula (2.22) defines a collinearity for all 3x3 (respectively 4x4) matrices. Proving this statement is not particularly difficult (see the formula described in the previous chapter on the parametric equation of a line). However, proving theorem 2.17 is much more complicated; see for example [Keré66] or [Fisc85] for the detailed proof.

Clearly, theorem 2.17 has the same importance for computer graphics as the existence of homogeneous coordinates, and for the same reasons: it becomes feasible to manage the collinearities numerically.

If the homogeneous coordinate system is generated out of a Cartesian one (following the method described in a previous chapter), the matrix representation gives also an easy way to decide whether a transformation is affine or not. Namely:

**Theorem 2.18.** If $T$ is a non-singular projective transformation of $\mathbb{P}E^2$ to $\mathbb{P}E^2$ and $\bar{T}$ is its matrix representation, $T$ is affine if and only if $\bar{T}$ is of the form:

$$\begin{bmatrix} t_{1,1} & t_{1,2} & t_{1,3} \\ t_{2,1} & t_{2,2} & t_{2,3} \\ 0 & 0 & \lambda \end{bmatrix} \tag{2.23}$$

where $\lambda$ is a non-zero real number.

In case of a transformation in the projective space, the corresponding form is

$$\begin{bmatrix} t_{1,1} & t_{1,2} & t_{1,3} & t_{1,4} \\ t_{2,1} & t_{2,2} & t_{2,3} & t_{2,4} \\ t_{3,1} & t_{3,2} & t_{3,3} & t_{3,4} \\ 0 & 0 & 0 & \lambda \end{bmatrix} \tag{2.24}$$

Theorem 2.18 is very easy to prove: spatial ideal points are uniquely characterised by the fact that their last coordinate value is zero, in other words, they are of the form $[(\alpha,\beta,\gamma,0)]$ where $\alpha,\beta$ and $\gamma$ are arbitrary real numbers with at least one of them being non–zero. The fact that the transformation $T$ is affine means therefore that

$$t_{4,1}\alpha + t_{4,2}\beta + t_{4,3}\gamma + t_{4,4}0 = 0 \tag{2.25}$$

for all possible non all-zero choices of $\alpha$, $\beta$ and $\gamma$. This means that $t_{4,1} = t_{4,2} = t_{4,3} = 0$ should hold; the value of $t_{4,4}$ must however be non-zero, to ensure non-singularity. ■

The reader may recognise the so–called *segment transformations* defined in GKS, GKS-3D, CGI etc. ([ISO85], [ISO88], [ISO88a]). However, the *modelling transformation* of PHIGS, PHIGS PLUS or PEX ([ISO89], [ISO89a], [ISO88b]) are not necessarily affine ones; indeed, the specification in these documents allows the user to give a general 4x4 matrix, without specifying any special features for the last row of it. This fact has severe algorithmic consequences on the so called *modelling clip* feature of these latter systems; more about that later. It is, however, strange that the GKSM specification in both the official GKS and the GKS-3D documents ([ISO85], [ISO88]) permit full 3x3 (resp. 4x4) matrices for segment transformations; this is clearly a mistake in the specification, as it would require the ability to handle a full, not necessarily affine transformation which is in contradiction with the rest of the specifications.

The different affine transformations used in computer graphics (rotations, scalings, shearings and translations) and their matrix representations are well described in a number of computer graphics textbooks (see all the already cited references) and it makes no particular sense to repeat these descriptions here[†]. In fact, the uniformity in description offered by the use of matrices was one of the reasons why homogeneous coordinates have become widespread in computer graphics even for 2D problems; it is a pity that even the newest textbooks on 3D graphics (like [Watt89]) do not explain why their use is mandatory when using projective transformations.

### 2.7.2. Viewing and Modelling Transformation

The main target of 3D systems is, after all, to render three dimensional objects on a two dimensional surface, which is the display screen or a plotter output. For that purpose, each such system has an internal mechanism which is usually called *viewing*. The most widespread approach to viewing is what is called the *synthetic camera model* in computer graphics literature and which is shown on figure 2.9. The idea is to project objects in a three dimensional frustrum onto the view plane; the frustrum (called the *view volume*) also serves as a clipping volume in space.

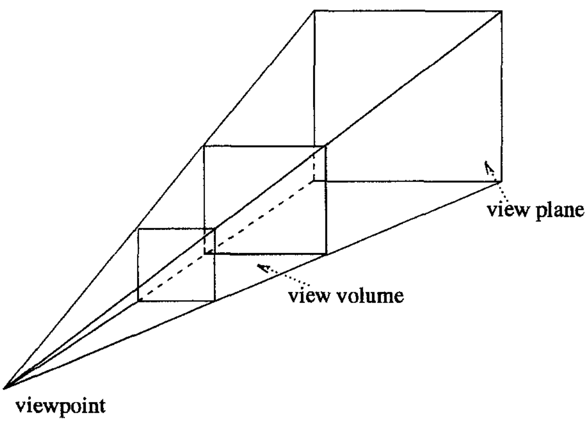



Figure 2.9.

The situation shown in figure 2.9 is usually referred to as "central projection" as opposed to the case where the the viewpoint is an ideal point, in which case the projection becomes a "parallel projection". In projective geometry terms, there is no difference between these two versions although, of course, the computational demands of a central projection are much greater than of a parallel one.

[†]Care should be taken, however, when using the different textbooks. Following the differences in conventions regarding operator–argument versus argument–operator notations in algebraic formulae, some of the textbooks might use vector–matrix multiplication rather than the convention adopted here. In such cases, the described matrices should be transposed to generate the appropriate version.

Early 3D graphics systems have effectively performed the projections as shown in the figure. This involved, however, two disagreeable consequences:

- clipping against the frustrum involves clipping against arbitrary planes in $I\!R^3$; this step is algorithmically demanding and
- performing the Hidden Line and/or Hidden Surface Removal in such a case is quite complicated as well.

As a result, in all newer systems (as well as in all 3D Graphics Standards or proposed Standards) the projection is done by first performing a special projective transformation in space (that is in $I\!PE^3$) which transforms the view volume onto a sub–cube of the unit cube and which transforms the viewpoint onto the ideal point of the $z$ axis of $I\!R^3$ (see figure 2.10). Having performed this transformation the projection itself becomes simply the projection onto the $x$–$y$ plane, the clipping against a view volume is reduced to a clipping against planes parallel to the base axes and, finally, the Hidden Line/Hidden Surface removal can be performed by applying appropriate algorithms which use exclusively the relative magnitude of the $z$ coordinate values of points (see for example [Mudu86] for an overview of such algorithms).

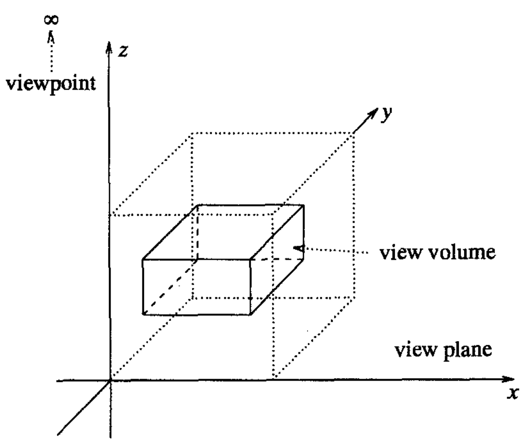



Figure 2.10.

The theorem about the existence and the uniqueness of a projective transformation (theorem 2.16) ensures that a mapping of the view volume onto the sub–cube of the unit cube uniquely exists. In general, this mapping will be non–affine (more precisely, the affinity of the transformation is equivalent to a parallel projection). There are several methods for the calculation of this transformation (also called *view transformation*); textbooks or tutorials like [Fole84], [Watt90], [Herm91], [Fole90] or others all describe the necessary steps and K. Singleton has also given a good and detailed overview of the GKS–3D/PHIGS cases ([Sing86]).

In this latter work, a description of the usual classification of the projections themselves is also given (oblique, isometric, 2–point perspective etc.); all these different mappings produce different visual effects on the screen, but their mathematical backgrounds are all the same. It is unnecessary to repeat all these formulae here; the reader is referred to the cited works (or alternative ones).

It has to be stressed that the ISO 3D documents do *not* state that the view transformation should be of the format described above. Not all projective transformations transform such a view frustrum onto a regular cube; there might be more complicated ones as well (eg the planes describing the view volume might not be parallel to the view plane). The user of a GKS–3D/PHIGS system has the possibility to specify any kind of 4×4 matrix as a view transformation and the system should be able to deal with it. The process described above is just offered as a utility (via the so–called utility functions of the specifications).

In PHIGS (and PHIGS PLUS), the viewing pipeline of the system includes yet another transformation which is called the *modelling transformation.* This transformation is primarily aimed at the use of convenient local coordinate systems for describing objects in 3D; by specifying the modelling transformation the appropriate coordinate transformation can be performed by the graphics system itself (in this sense, this transformation is the generalisation of the so called normalisation transformation of the GKS and GKS–3D specifications, [ISO85], [ISO88a]). However, to achieve some special effects (like the image of a projection modelled by three dimensional objects) this transformation is to be specified by a full 4×4 matrix, allowing therefore a general, not necessarily affine transformation.

The view transformation and the modelling transformation are the two, not necessarily affine, transformations arising in 3D computer graphics systems. In what follows, no special attention will be paid to their usual use in practice and their format; as described above, the 3D systems at hand must be able to handle these transformations in all their generality.

### 2.7.3. Description of Collinearities Based on the Straight Model

The straight model of the projective plane and/or space also permits the visual representation of the effects of a projective transformation. This is important because just as the straight model creates a strong link between $\mathbb{P}E^2$ and $\mathbb{R}^3$ (respectively between $\mathbb{P}E^3$ and $\mathbb{R}^4$), the visualisation to be described creates a link between projective transformation and the linear transformation of $\mathbb{R}^3$ and $\mathbb{R}^4$ respectively. This is of no particular interest in traditional projective geometry and it is therefore never described in the previously cited projective geometry textbooks; however, for the purposes of computer graphics where the linearity of some traditional problems has importance, its use has proven to be extremely useful (see [Herm87], [Herm88], [Herm91], [Hübl90]).

As in the case of the projective points and lines of $\mathbb{P}E^2$, projective transformations described acting on $\mathbb{P}E^2$ can be visualised easily. It is worth recalling (see also figure 2.11) that in the straight model $\mathbb{P}E^2$ is identified with the points in $\mathbb{R}^3$ on

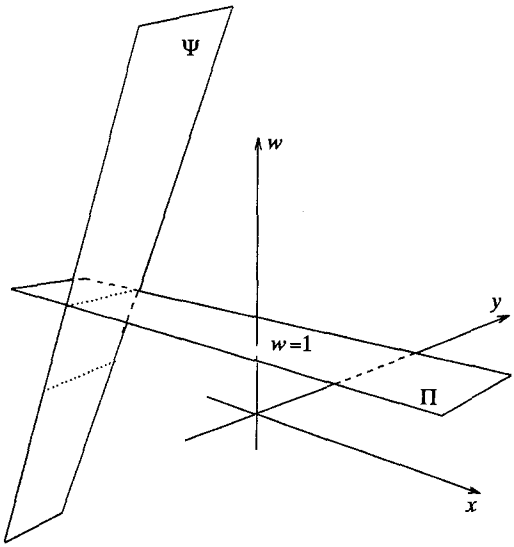



Figure 2.11.

the plane $w = 1$ (denoted by $\Pi$ on figure 2.11). A projective transformation $T$ is represented by a matrix $\bar{T}$. This matrix $\bar{T}$ also generates a traditional linear transformation $T'$ in $I\!R^3$ using the matrix-vector multiplication. This transformation will map $\Pi$ onto another plane of $I\!R^3$ which has, in general, an arbitrary position in space. This plane is denoted by $\Psi$ on the figure. (In fact, $\Psi$ can also be considered as an alternative straight model of $I\!PE^2$). To get back to the more usual $w = 1$ model, $\Psi$ has to be projected centrally (through the origin) back onto $\Pi$; this is, in fact, the so-called projective division.

What can be said therefore, in view of figure 2.11, is as follows. A transformation $T$ of $I\!PE^2$ can be modelled by a *two-stage process*. First, $\Pi$ is mapped by the *linear* transformation $T'$ (of $I\!R^3$) onto $\Psi$ and, secondly, the resulting plane $\Psi$ is projected back onto $\Pi$ by a central projection through the origin. The first step might be called the *linear part* of the transformation while the second step is traditionally denoted as the projective division. This simple fact has far-reaching practical consequences; in fact, the whole of the next chapter will be based, in some sense, on this observation.

Clearly, $T$ is affine if and only if $\Psi$ is parallel to $\Pi$ (on figure 2.11). Furthermore, the usual description of the affine transformation (that is where $T_{4,4} = 1$) results in $\Pi$=$\Psi$.

In the case of $I\!PE^3$, $\Pi$ and $\Psi$ are three dimensional subspaces of $I\!R^4$, also called hyperspaces. With this difference, the whole description remains valid for $I\!PE^3$ as well.

## 2.8. Division Ratio and Cross Ratio

As stated before, the notion of line segments has no meaning in projective geometry any more. Likewise, the distance between two points also becomes an unclear notion; indeed, the distance is usually defined in terms of the generated line segment. Furthermore, even if the distance could be defined at least in restricted cases (eg by excluding ideal points), the projective transformation does not retain these values. It is to find some kind of a stable numerical value that the notion of double ratio has been introduced in projective geometry; in a number of cases, its use is necessary to prove some of the important statements of the theory. Furthermore, as will be presented later, the double ratio can also be a very useful tool for the purposes of computer graphics, that is the reason why it is presented here.

The division ratio and the double ratio will be defined in terms of affine and ideal points, that is making use of the construction which has led to a projective plane and/or projective space. A mathematically "pure" definition (that is based on the axiomatic system only) would also be possible, but it would be more abstract and more complicated as well (see eg [Coxe74]). For the purposes of computer graphics, the more "pragmatic" approach is acceptable.

At first, the notion of *division ratio* has to be defined as follows.

**Definition 2.7.** Let $A$, $B$ and $C$ be three different collinear affine points on $\mathbb{P}E^2$ or $\mathbb{P}E^3$. The *division ratio* of the points $A$, $B$ and $C$ (denoted by $(ABC)$ ) is defined to be the following (non-zero) real number:

$$(ABC) = \frac{\overline{AC}}{\overline{CB}} \tag{2.26}$$

where $\overline{AC}$ and $\overline{CB}$ mean the *directed* Euclidean distances of the two points (that is $\overline{AC} = -\overline{CA}$).

If the point $C$ tends to infinity, the limit of the corresponding division ratio (with the points $A$ and $B$ remaining fixed) will be $-1$; consequently, it seems to be feasible to extend (2.26) to the case when the point $C$ is an ideal point, namely let

$$(ABC) = -1 \tag{2.27}$$

in this case.

With the help of the division ratio the *double ratio* (also called sometimes the *cross ratio*) of four collinear points may be defined as follows.

**Definition 2.8.** Let $A$, $B$, $C$ and $D$ be four different collinear points of $\mathbb{P}E^2$ or $\mathbb{P}E^3$ not all four being ideal. The *double ratio* of the four points (denoted by $(ABCD)$ ) is the real number defined by:

$$(ABCD) = \frac{(ABC)}{(ABD)} \tag{2.28}$$

Clearly, when all four points are affine (i.e. none of them is ideal), the double ratio may also be expressed with the help of directed distances, namely:

$$(ABCD) = \frac{\overline{AC}\,\overline{DB}}{\overline{CB}\,\overline{AD}} \tag{2.29}$$

The double ratio has a number of remarkable properties. Some of them will be cited here, which are necessary for later purposes; the corresponding proofs may be found for example in Penna and Patterson ([Penn86]), in Fischer ([Fisc85]) or in any other standard textbook on projective geometry†.

First of all, the double ratio is a one–to–one mapping of the points of the line and the (non-zero) real numbers. Namely, the following is true:

**Theorem 2.19.** Given three different collinear points on the projective plane, denoted by $A$, $B$ and $C$, if $x$ is an arbitrary non-zero real number, then there exists one and only one point $D$ on the line determined by $A$, $B$ and $C$, for which the following equation holds:

$$(ABCD) = x \tag{2.30}$$

The second property has a particular importance for computer graphics (and, in fact, it is one of the most important results in projective geometry). The description of this property requires first a definition:

**Definition 2.9.** The projective invariance of the double ratio is defined as follows. If $A$, $B$, $C$ and $D$ are four different arbitrarily chosen collinear points of the plane (or space) and $T$ is an arbitrary projective transformation, then the following property should be valid:

$$(ABCD) = (T(A)T(B)T(C)T(D)) \tag{2.31}$$

The projective invariance of the division ratio means that for each three points $A$,$B$ and $C$ the relation

$$(ABC) = (T(A)T(B)T(C)) \tag{2.32}$$

holds.

The affine invariance of the double ratio and the division ratio can be defined similarly; in this case $T$ should be affine.

†One should be careful again, however, when consulting the literature; in some cases, following different local traditions, the definition of the double ratio may slightly differ from the one given here (e.g. by an additive constant, the order of the directed distances in the formulae etc.).

With this definition at hand, the following theorem holds:

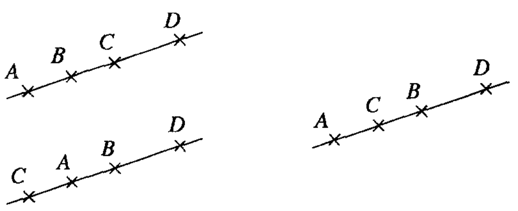
**Theorem 2.20.** The double ratio is projective invariant, the division ratio is affine invariant.

See for example [Fisc85] or [Penn86] for a detailed proof.

Theorem 2.20 also shows that the value of the double ratio is independent of the construction of $\mathbb{P}E^2/\mathbb{P}E^3$. Although this has not been explicitly stated up to now, the Euclidean analogy works here perfectly, and the change of homogeneous coordinate system on $\mathbb{P}E^2$ or $\mathbb{P}E^3$ can be described by a matrix–vector multiplication, using, of course, a homogeneous matrix. In other words, this is analogous to the use of a projective transformation which, according to theorem 2.20, leaves the value of the double ratio unchanged.

It should be remarked that a stronger statement regarding invariance of the division ratio is not true. That is, the the division ratio is *not* projective invariant.

Penna and Patterson in [Penn86] give some methods to calculate the exact value of the double ratio for four given collinear points in $\mathbb{P}E^2$. Without going into details, the idea is to project the points onto the main coordinate axes (which can be done, in fact, by replacing one of the coordinate values by 0) and calculate the double ratio of the resulting four points. Taking into account that the projection keeps the value of the double ratio, this is clearly a valid approach.




a) $(ABCD) > 0$ b) $(ABCD) < 0$

Figure 2.12.

In some cases, however, not the exact value but only the *sign* of the double ratio is of interest. Indeed, it is easy to prove (using formula (2.29), see also figure 2.12) that if for the points $A$, $B$, $C$ and $D$ the value of $(ABCD)$ is negative, the (affine) line segments $AB$ and $CD$ "cut" (overlap) one another while that is not the case if the value of $(ABCD)$ is positive (see figure 2.12). The importance of this fact is that as an arbitrary projective transformation keeps the value of the double ratio, this "segment cutting" property is invariant for projective transformations.

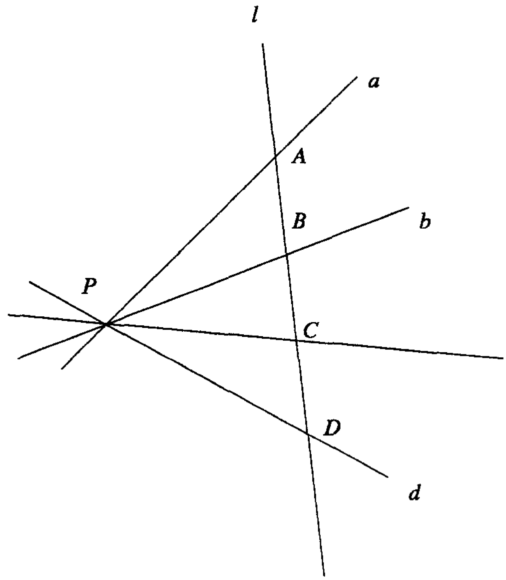



Figure 2.13.

Theorem 2.20 permits the assignment of a double ratio to four lines of $\mathbb{P}E^2$ which have one common intersection point (such a configuration will be referred to as a "bunch of lines"). Figure 2.13 shows how this can be done. If $a$, $b$, $c$ and $d$ are the four lines intersecting at $P$, let us take any line $l$ *not* containing $P$. If $l \wedge a = A$, $l \wedge b = B$ etc., the following definition can be given:

$$(abcd) = (ABCD) \tag{2.33}$$

The fact that this assignment can be done is by itself is not very surprising; indeed, a bunch of lines (that is four lines intersecting at one point) is the dual form of four points lying on the same line. The definition is not dependent upon the choice of the line $l$: if another line, say $l'$ were taken, the central projection with $P$ as a centre would map $l$ onto $l'$, mapping the corresponding intersection points onto one another; consequently, according to theorem 2.20, the definition in (2.33) is indeed correct. This also means that the "segment cutting" property has also its counterpart for such a configuration, although it should rather be called "domain cutting" in this case.

Formula (2.33) also fills a "hole" in the definition of the double ratio. If four collinear points $A,B,C$ and $D$ are all ideal, there has been no definition given up to now for the value of $(ABCD)$. However, four ideal points generate four lines by choosing an arbitrary point $P$ on $\mathbb{P}E^2$; formula (2.33) gives then the value of $(ABCD)$. Using the theorem on the existence of a projective transformation (theorem 2.16), if a different point $Q$ is chosen instead of $P$, there exists a projective transformation which would transform the corresponding lines onto one another; as the double ratio is projective invariant for lines also, this definition holds.

The invariance of the double ratio provides a method to prove a number of so–called permutation formulae on double ratio. Two of them are as follows:

**Theorem 2.21.** If $A,B,C$ and $D$ are four different collinear points, then the following equalities hold:

$$(ABCD)(ABDC) = 1 \tag{2.34}$$

and

$$(ABCD) = (CDAB) \tag{2.35}$$

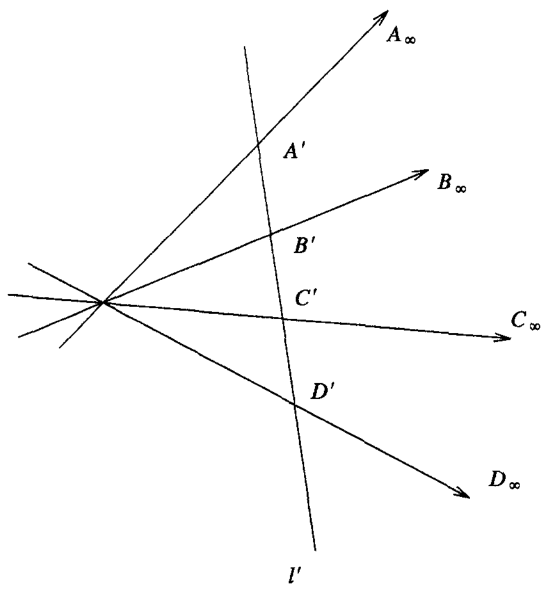



Figure 2.14.

The proof for these formulae is very simple: it is enough to prove the case where none of the points $A,B,C$ or $D$ is ideal. If this is not the case, it is possible to choose four appropriate affine points $A',B',C'$ and $D'$ and a projective transformation (actually a central projection) which would map one set of points onto the other. Figure 2.14 shows the case when all four points are ideal (and projected onto the line $l'$) and figure 2.15 is the case when only one of the points is ideal (and projected onto the line $l'$ again). The points being collinear, there are no other alternatives. If the equations are true for the points $A'$, $B'$, $C'$ and $D'$, they are also true for $A$, $B$, $C$ and

$D$ because of the invariance of the double ratio. Finally, to prove that the equations are true for the purely affine case is simply a matter of algebraic exercise, based on the definition of the double ratio. ■

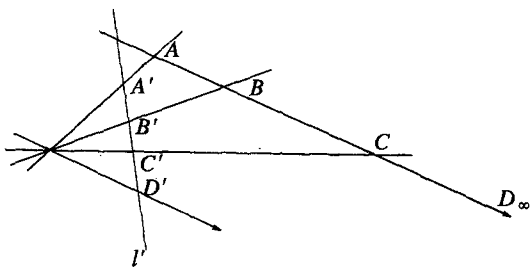



Figure 2.15.

For $I\!PE^3$, a double ratio can be assigned to each four *coplanar* lines intersecting in one point. Furthermore, if the planes $\Pi$, $\Psi$, $\Phi$ and $\Theta$ are such that they intersect in one common line, the value of the double ratio $(\Pi\Psi\Phi\Theta)$ can be defined as well, using formula (2.33); this also means that the "domain cutting" property is also valid in this case.

In [Kram89] G. Krammer introduced the notion of *conic sectors*, which is an interesting application of the double ratio. In what follows, conic sectors will be defined for $I\!PE^2$; in case of $I\!PE^3$, planes should be taken instead of lines to arrive at the same definitions.

If two lines, say $l$ and $n$, are given on $I\!PE^2$, these lines will cut $I\!PE^2$ into two disjoint subareas. The definition is as follows.

**Definition 2.10.** If $P,Q \in I\!PE^2$ are given, let $X$ (resp $Y$) denote the intersection points $(P \vee Q) \wedge l$ (resp. $(P \vee Q) \wedge n$). In case $X = Y$, the points $P$ and $Q$ are said to be in the same conic sector. If $X \neq Y$, $P$ and $Q$ are said to be in the same conic sector if and only if the following inequality holds:

$$(PQXY) > 0 \tag{2.36}$$

See also figure 2.16. There might be some special cases for conic sectors. If $l$ and $n$ are parallel in the Euclidean sense (that is their intersection point is an ideal point), the corresponding conic sector is shown in figure 2.17a). If, finally, $l$ is affine but $n$ is the ideal line, the two conic sectors are the two half-planes! (see figure 2.17b).

The interesting thing about conic sectors is the fact that conic sectors are transformed into conic sectors by projective transformations. Indeed, projective transformations map intersection points onto intersection points; that is, if $T$ is the transformation in use then, using the notation of figure 2.16:

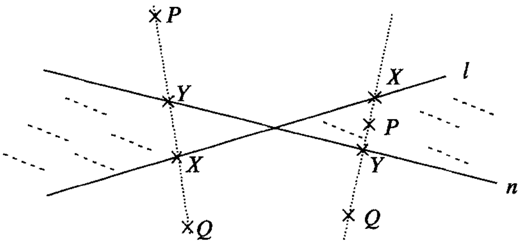



Figure 2.16.

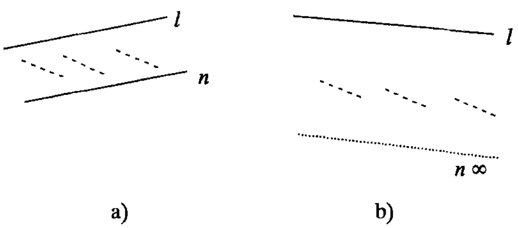



Figure 2.17.

$$\begin{aligned}(T(P)\vee T(Q))\wedge T(l) &= T(X)\\ (T(P)\vee T(Q))\wedge T(n) &= T(Y)\end{aligned} \tag{2.37}$$

furthermore, the projective transformation also keeps the double ratio, that is

$$(PQXY) = (T(P)T(Q)T(X)T(Y)) \tag{2.38}$$

in other words, the sign of the double ratio will also remain unchanged.

In Euclidean geometry, convex polygons or convex polyhedra can be described by the intersection of a finite number of half planes/spaces. Conic sectors may be considered as the generalisation for the projective case of half–spaces (or half–planes); in other words, one can speak about projective polygons as the intersection of a finite number of conic sectors. The importance of this approach will become clear later.

## 2.9. Projective Theory of Conics

### 2.9.1. Introduction

Besides line segments and polygons, conics also occur frequently in computer graphics. They are used in geometric design, they are frequently implemented as GDPs (Generalized Drawing Primitives) in various graphics packages and some are included in the basic set of output primitives of ISO documents (e.g. CGI, [ISO88a]). Among the three main classes of conics, namely ellipses, parabolae and hyperbolae, the use of ellipses (first of all circles) is the most widespread. Circles and circular arcs are used in business graphics for charts, in mechanical engineering for rounded corners, for holes etc. Circular arcs may also be used to interpolate curves (see for example Sabin in [Sabi77]). Although the role of parabolae and hyperbolae is not so important, they cannot be ignored either. On the one hand they do appear in practical applications (for example there are proposals to use parabolic arcs for curve approximation like the so-called double-quadratic curves in Várady in [Vára84] or [Vára85]) but, principally, these curves appear automatically when distorting an ellipse with a projective mapping.

Projective geometry gives a unified framework for handling all kinds of conics. This might sound surprising at first glance, as projective geometry is considered to be a theory primarily concerned with lines and their behaviour. However, if one thinks of the well–known fact that the conics appear as the planar intersections of cones, which is very much like the figure of a central projection, it becomes more plausible that projective geometry has this descriptive power for conics.

What is the real problem as far as computer graphics is concerned, in handling these curves? Mathematically, (planar) conics are described by a second order polynomial of the form:

$$a_{1,1}x_1^2 + a_{2,2}x_2^2 + 2a_{1,2}x_1x_2 + 2a_{1,3}x_1 + 2a_{2,3}x_2 + a_{3,3} = 0 \tag{2.39}$$

While this formula is appropriate to perform all calculations which are necessary in a modelling system (see for example Fraux and Pratt [Faux79]) it is inadequate to *draw* the corresponding conic. Indeed, practically all graphics devices available today are designed to render (in hardware/firmware) line segments; in other words, the "ideal" mathematical curve must be approximated by an appropriate polyline or polygon. To achieve a reasonable appearance, this approximation must be quite dense; for example the number of approximation points to render a circle properly must be at least 100, but an approximation with 360 points (that is one point for each degree) may also be necessary.

It is not an easy task to generate these points properly. Appropriate approximation formulae or equations are necessary; some examples will be presented later. Some of these formulae (especially those describing ellipses) are already known to the graphics community, whereas some others are relatively unknown. Furthermore, having these formulae at hand does not solve all problems. An implementor has to give an answer to the following question: where in the graphics output pipeline is the approximation effectively performed?

The approach usually chosen is to approximate the curve with a polygon/polyline before performing a transformation in the pipeline. Most of the formulae described in different textbooks and used in practice are not projective invariant, that is the data generating these formulae change their geometrical nature when applying a projective transformation. Therefore, the curves are approximated beforehand, the resulting polygons/polylines are transformed and rendered following already well established methods.

There are, however, some problems with this approach. First of all, there is a loss in speed and storage. As mentioned already, the number of generated points tends to be relatively large; all these points have to be transformed, that is a matrix–vector multiplication has to be applied and, in the case of a projective and non–affine transformation, an additional projective division must also be performed. By applying some alternative methods presented later (primarily in chapter 5), a speed improvement of at least 25% can be achieved. This figure might not seem very impressive at a first glance but one should never forget that computer graphics is (ideally) interested in real–time effects where a 20%-25% improvement might be of real significance.

A widespread approach to overcome this difficulty is to use the so–called rational B–splines to describe conics. Second order rational B–splines can describe any kind of conic and this is done in a more or less projective–invariant manner (see eg [Faux79], [Till83], [Pieg87], [Fari88]). Beside the fact that B–Splines are computationally expensive (see all the calculation formulae in [Bart85] or [Bart87]), this approach leads still to another common problem: the quality of the approximation.

Speed is not the only issue (and having all these super–fast computers invading the market, this argument might be less and less important). However, when approximating for example a circle with 360 points, one gets a fairly regular geometrical ordering of the points which, if displayed directly, will produce an acceptably smooth shape. However, if a transformation is applied against this set of points, this "regularity" will be lost. Some of the line segments will become much longer than others; in these areas the resulting polyline will have a "jagged" effect whereas on some other parts of the curve the density of the points will be unnecessarily high. It is very difficult to keep track of these distortions which may be, in the case of a more complicated projective transformation, very noticeable. The only way of reducing this effect is to postpone the approximation step as "far" as possible and to produce the resulting polyline after the transformations instead of prior to it.

The real difficulty with this approach is the fact that a non–affine projective transformation will "destroy" a number of geometrical characteristics of the points. As an example remember that the centre of an ellipse might not be the centre any more; furthermore, the image of an ellipse is not even an ellipse in some cases; it may become a hyperbola or a parabola. Consequently, a thorough investigation of the nature of conics is necessary, using the tools of projective geometry. This is why this part of the theory has also been included in the present study.

### 2.9.2. General Theory of Conics

#### 2.9.2.1. The 2D Case

As usual, planar conics will be treated first; some of the ideas will then be generalised for space as well.

In a Euclidean environment, a conic is described by the equation (2.39). By defining the symmetric matrix $A = (a_{i,j})_{i,j=1}^{3}$ and by using homogeneous coordinates instead of Euclidean ones (with the usual identification mechanism), the equation has its counterpart for projective environments as well, namely:

$$\sum_{i,j=1}^{3} a_{i,j} x_i x_j = 0 \qquad (2.40)$$

Finally, formula (2.40) can be abbreviated by the so called bilinear form, that is:

$$xAx = 0 \qquad (2.41)$$

The notation of formula (2.41) will be used throughout the whole chapter. In the whole section $A$ will be considered to be a non-singular matrix, that is $det(A) \neq 0$ (some of the theorems presented later are not valid for singular cases; on the other hand, the corresponding "curves" in this case are lines, points or just the empty set).

In fact, (2.41) can be used in a somewhat more general way to define the notion of *conjugate points*. This definition is as follows:

**Definition 2.11.** The points $x,y \in \mathbb{P}R^3$ on the projective plane are said to be conjugate points with respect to the conic represented by the symmetric matrix $A$ if and only if the following equation holds:

$$xAy = 0 \qquad (2.42)$$

($A$ being symmetric, $xAy = yA^T x = yAx$ holds). We could also say that the points of the curve may be characterised by the fact that they are auto-conjugate.

The notion of conjugate points has a number of nice properties. Indeed, the following facts are true (their proofs may be deduced from the definitions or they may be found in the textbooks cited above):

**Theorem 2.22.** If $x \in \mathbb{P}R^3$ is a fixed point, the set of all points $y \in \mathbb{P}R^3$ which are conjugate to $x$ with respect to a conic represented by the symmetric matrix $A$ form a line of the projective plane. This line is called the *polar* of $x$; it may be represented by the homogeneous vector $Ax$.

**Theorem 2.23.** If $l$ is a line in the projective plane, then there is one and only one point whose polar with respect to a conic represented by

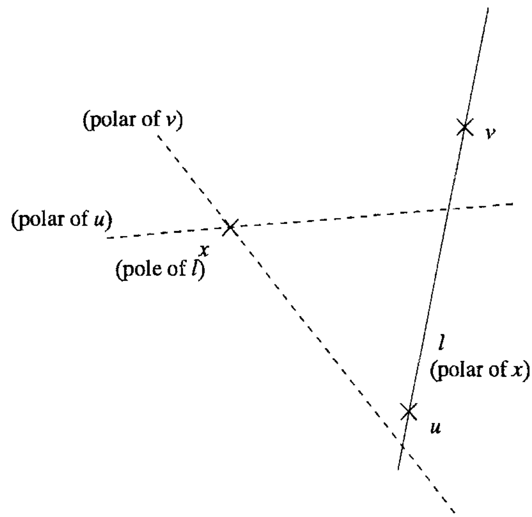



Figure 2.18.

the symmetric matrix $A$ is $l$; this point is called the *pole* of $l$. The pole may also be characterised as follows: it is the (unique) intersection point of all the polars generated by the points on $l$ (see also figure 2.18).

**Theorem 2.24.** If $x \in \mathbb{P}\mathbb{R}^3$ is the homogeneous vector of a point on the projective plane, then $x$ belongs to its own polar if and only if $x$ is a point of the conic itself. In this case, the polar of $x$ will be tangential to the conic at the point $x$ and the homogeneous representation of this tangential line is $Ax$.

**Definition 2.12.** The pole of the ideal line is called the *centre* of the curve; for ellipses and hyperbolae, this coincides with the traditional, Euclidean definition of the centre of these curves.

Two lines $l_1$ and $l_2$ are said to form a *conjugate pair of lines* if the pole of $l_1$ belongs to $l_2$ and, conversely, the pole of $l_2$ belongs to $l_1$. One may speak of a pair of *conjugate chords* as well as of a pair of *conjugate diameters*, denoting a pair of conjugate lines which are chords (resp. diameters) of the conic (diameter is a chord containing the centre).

All these definitions are, unfortunately, rather abstract and a certain time is needed to get used to them and to get an intuitive feeling as far as their geometrical meaning is concerned. Figure 2.19 shows an example which might help in using these definitions. The line $l$ has two intersection points with the conic, $P$ and $Q$. The polars of these points are the two tangents $l_1$ and $l_2$ respectively. In view of what has been said before, the intersection point of these lines, that is $R$, is the pole of the

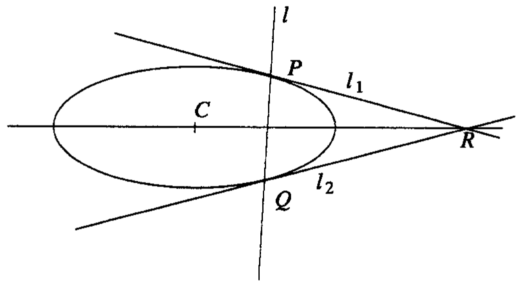



Figure 2.19.

line $l$. This procedure is the usual way of generating the pole of a line, provided the line has two intersection points with the curve (which is not always the case). It can also be remarked that if the centre of the curve is denoted by $C$ (like in the figure), the line $R \vee C$ will intersect the line segment $PQ$ in its middle point. Also, the pole of the lin $R \vee C$ will be on $l$, that is a pair of conjugate lines have been created (see for example [Keré66] for a proof of these features).

The importance of these definitions becomes clearer when the behaviour of conics in relationship to projective transformations is examined. If the matrix of the transformation is denoted by $T$, and if a conic is represented by the symmetric matrix $A$, then for all $x,y \in \mathbb{P}\mathbb{R}^3$:

$$xAy = x^T(Ay) = \; = x^T(T^{-1}T)^T(AT^{-1}Ty) = \tag{2.43}$$
$$x^TT^T(T^{-1})^T(AT^{-1}Ty) = (Tx)^T((T^{-1})^TAT^{-1})(Ty)$$

Now, if the notation

$$T(A) = (T^{-1})^TA(T^{-1}) \tag{2.44}$$

is applied, then (2.44) can be simplified to $(Tx)T(A)(Ty)$. In other words, the image of a conic under the effect of a projective transformation remains a conic and, furthermore, formula (2.44) gives an easy way to calculate the matrix of the image. Also, the property of conjugation is projective invariant. The pole–polar relationship also remains valid across the transformation. However, the image of a centre is *not* necessarily a centre: although it is true that the image of the centre will still be the pole of the image of the ideal line, it is not necessarily the case that the image of the ideal line will still be the ideal line (it is however affine invariant).

The regularity of the matrix $A$ has an interesting consequence, which is as follows.

**Theorem 2.25.** If $p,q \in \mathbb{P}\mathbb{R}^3$, $p,q \neq 0$ are such that

$$pAp = qAq = pAq = 0$$

then $p = q$ (where the equality is meant in homogeneous sense).

Let $u = Ap$ and $v = Aq$. The vectors $u$ and $v$ are usual three dimensional vectors, that is $u,v \in \mathbb{R}^3$. None of them is the zero vector, as $det(A) \neq 0$. If $u$ and $v$ are equal in the homogeneous sense, then $p = q$ follows (again in the homogeneous sense). If this is not the case, then, from the premises of the theorem, $p$ as a regular three dimensional vector is perpendicular to both $u$ and $v$. In other words, $p$ is perpendicular to the plane $u \vee v$. However, the same is true for $q$ and this is possible only if there exists a $\lambda \in \mathbb{R}$ such that $p = \lambda q$. ■

The mutual relationship of a conic and a line is of particular interest. Namely:

**Theorem 2.26.** The number of intersections of a line and a conic may be 0, 1 or 2.

Although this fact is well-known in projective geometry, it is worth presenting the proof of this theorem here as well. The reason is that the proof gives an effective way of calculating the (possible) intersection points and this is a useful feature in what follows. It will also be necessary to make use of the result of theorem 2.25.

The line to intersect with can be described by (see (2.18)):

$$p \vee q = \{ \lambda p + \mu q \} \qquad (2.45)$$

$$\lambda \neq 0 \text{ or } \mu \neq 0$$

where $p$ and $q$ are two points on the line. Two values for $\lambda$ and $\mu$ are searched for which

$$(\lambda p + \mu q) A (\lambda p + \mu q) = 0 \qquad (2.46)$$

holds. In fact, because of the homogeneous nature of the formula, one is not interested in the exact values of $\lambda$ and $\mu$ but only in their relative ratio $\lambda/\mu$. Equation (2.46) can be also rewritten by:

$$pAp\lambda^2 + 2pAq\lambda\mu + qAq\mu^2 = 0 \qquad (2.47)$$

First of all, $pAp$ and $qAq$ cannot be both zero. Indeed, this would mean

$$2pAq\lambda\mu = 0 \qquad (2.48)$$

for all possible choices of $\lambda$ and $\mu$, that is $pAq = 0$ would also hold; however, according to theorem 2.25, this is not possible ($p$ and $q$ are considered to be different points of $\mathbb{P}\mathbb{E}^2$). If $pAp \neq 0$, $\mu \neq 0$ may be considered; if this were not the case, then $\lambda = 0$ would also hold, which is impossible. Similarly, if $qAq \neq 0$ then $\lambda \neq 0$. Let us consider the first case; that is the equation can be divided by $\mu^2$ to get

$$pAp(\lambda/\mu)^2 + 2pAq(\lambda/\mu) + qAq = 0 \qquad (2.49)$$

clearly, this equation (in $\lambda/\mu$) may have 0, 1 or 2 solutions; by solving it one also gets an explicit value for the (possible) intersection point(s). ■

The relationship between lines and conics has a very important consequence, as the theorem can be applied to a special case to get a simple means of classification for conics. Namely:

> **Theorem 2.27.** The number of ideal points belonging to a conic may be 0, 1 or 2. (The set of ideal points being the ideal line, this is just the special case of the previous statement). If this number is 0, the curve is an ellipse (or a circle); if it is 1, the curve is a parabola with the axis determining its ideal point; and if it is 2, the curve is a hyperbola, with the two asymptotes determining the two ideal points.

Theorem 2.27 (and the previously cited features) are of particular importance for computer graphics. Some of the consequences are:

- The tangent of a curve remains a tangent and a chord remains a chord after transformation: the intersection points of lines and curves are auto-conjugate points and the conjugation is a projective invariant property.
- Each class of conics is *affine invariant*. In other words, the affine image of an ellipse will be an ellipse, the affine image of a parabola will be a parabola etc. In the case of parabolae for example, the ideal line is tangential to the curve; the image of a tangent being still a tangent and the image of the ideal line being still the ideal line, the image curve has only one ideal point. In other words, the image of the curve has still one ideal point only, which means, according to theorem 2.27, that the image is a parabola. The same reasoning holds for ellipses and hyperbolae as well.
- By using straightforward and simple calculations it is easy to decide from the matrix of a curve which class the curve belongs to: the way theorem 2.26 was proven gives also a way to calculate the intersection points (if any), and also to calculate their number. All what is required, is to use the homogeneous coordinate values of two ideal points (see 2.6.1 for $\mathbb{P}E^2$ and 2.6.2 for $\mathbb{P}E^3$).

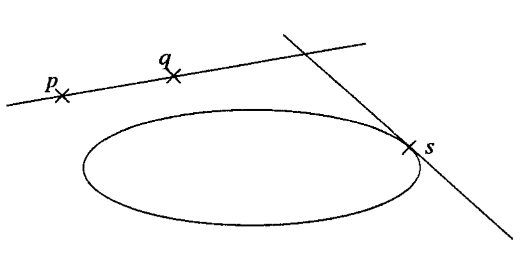



Figure 2.20.

Let this section be concluded by yet another calculation formula for planar conics which will be useful later. The task is as follows: if $p,q \in I\!PR^3$ are two points, $A$ is a symmetric matrix representing a conic and, furthermore, $s \in I\!PR^3$ is a point on the conic, compute the intersection of the tangent at $s$ and $p \vee q$ (see figure 2.20). This could be done reusing already known formulae but an alternative (and computationally more attractive) method is as follows. Once again, appropriate $\lambda$ and $\mu$ numbers are to be found so that:

$$(\lambda p + \mu q)As = 0 \tag{2.50}$$

taking into account that $As$ gives the homogeneous representation of the tangential line at $s$. That is:

$$\lambda pAs + \mu qAs = 0 \tag{2.51}$$

Again, if $pAs \neq 0$ then $\mu \neq 0$, that is we can divide; the result is:

$$\lambda / \mu = -qAs / pAs \tag{2.52}$$

### 2.9.2.2. The 3D Case

The notion of conics may be generalised for projective spaces as well; the only difference is that the symmetric matrix in use should be 4×4 instead of 3×3. These conics are the so-called quadratic surfaces (hyperboloids, paraboloids, hyperbolic paraboloids etc.). Their classification is much more complicated than in the case of planar curves; however, they form again a class of surfaces which is invariant to projective transformations (the way theorem (2.44) has been deduced was independent of dimensions). Many of these quadratic surfaces are rarely used directly in computer graphics, except some of the symmetric rotation surfaces. In such cases, the rational B-spline formulation for these surfaces is the widely accepted approach (see again [Till83], [Pieg87], [Fari88] and also [Klem89]).

As quadratic surfaces appear in very special cases only, no further investigation will be presented here. Instead, this section will concentrate on what will happen to planar quadratic curves in a 3D environment; taking into account their usefulness in practice, it is worth examining this special case more thoroughly.

One way of handling planar conics in space is to find a vector representation of a conic curve, that is an equation which describes the points of the curve as a function of some vectors and some additional real parameters. Such a formula would be useful if it were at least affine invariant, that is if the transformation of the vectors of the equation were enough to describe the transformed curve. Such an affine invariant formula can be found for all three classes of the curves; while the formula describing an ellipse has been known for quite a long time, the corresponding formulae for parabolae and hyperbolae had to be reconstructed from different mathematical bits and pieces (these formulae were never of a real interest to mathematicians, that is why they are not usually presented anywhere). This has been done in [Herm89a] and will also be presented in a later chapter.

To perform some computations, however, a more complicated approach is also necessary which is as follows. The intersection of a plane and a quadratic surface leads to a planar conic on the plane. If the plane happens to be the plane $x_3 = 0$, this can be easily seen by just putting a 0 to all relevant places of the equation of the surface; the result is a second order equation for the remaining coordinate values. If the plane is of a general position, it can always be transformed into the plane $x_3 = 0$ by using an orthogonal transformation. These intersection curves are from now on the main focus of interest. Also, it is easy to associate a quadratic surface with a planar conic: one has to construct a *generalised cylinder* (it might also be called a sweep surface). This means that the curve should be moved along a line not contained by the plane of the conic (see figure 2.21). In the simplest case, when the conic lies in the $x-y$ plane, it is also very simple to give the equation of such a surface. If $A$ is a 3×3 matrix then the matrix describing the corresponding surface may be:

$$A_c = \begin{bmatrix} a_{1,1} & a_{1,2} & 0 & a_{1,3} \\ a_{2,1} & a_{2,2} & 0 & a_{2,3} \\ 0 & 0 & 0 & 0 \\ a_{3,1} & a_{3,2} & 0 & a_{3,3} \end{bmatrix} \tag{2.53}$$

If the plane is not the $x-y$ plane, an affine transformation and 2.21 should be applied; concrete examples will be given in chapter 4 (see also [Herm89a]). Notice should be taken of the fact that the matrix $A_c$ is singular; however, its rank is 3 (that is it contains a 3×3 non-singular submatrix). In fact, it can be shown that in case of 3D, if the matrix of a quadratic surface is singular but its rank is 3, it is either the matrix of a (generalised) cylinder or that of a cone (there is no difference in a projective sense between a cylinder and a cone: the cylinder is a cone whose focal point is ideal). For the proof of this theorem the interested reader should consult for example [Keré66].

Using (2.53), a three dimensional surface can be assigned to each planar conic. Using then (2.44), the image of this surface under the effect of a transformation can be described. As a next step some characteristic data of the *intersection of the surface and a plane* should be calculated; indeed, what a computer graphics system is really interested in is not the whole surface but only the planar cut of it. To use all the formulae of the previous section, the following are needed:

- If $p,q \in \mathbb{P}\mathbb{R}^4$ are two points *in the plane* $\Pi$ and $A$ is a 4×4 symmetric matrix representing a quadratic surface, compute the number of intersection points and the eventual intersection points themselves of $p \vee q$ and the (planar section of) the surface.

The same calculations as the one presented for the two dimensional case can be adapted to 3D as well. Care should be taken, however, that in this case it is possible that the intersection is a line. Theorem 2.25 is not valid in $\mathbb{P}\mathbb{E}^3$ any more (the arguments used to prove it were very much bound to the nature of $\mathbb{E}^3$); in other words, it is generally possible that a quadratic surface would contain a whole line (see

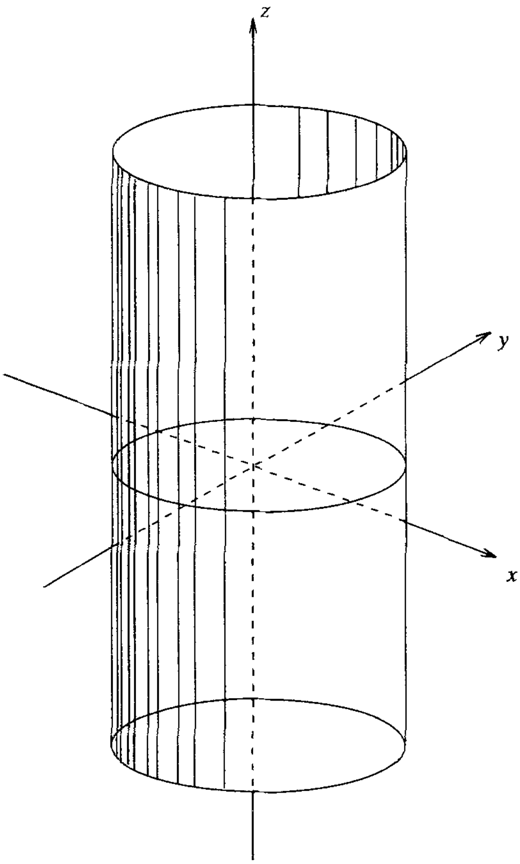



Figure 2.21.

figure 2.21 or, to choose a regular case, the well known saddle surface). But this situation is equivalent to the fact that $pAp = qAq = pAq = 0$, and this can be checked easily. In particular, if $\Pi$ is the plane containing the original conic and $A$ is the matrix of the transformed cylinder, by taking two ideal points of $A(\Pi)$, the exact classification *of the planar section* can be done.

- If $p,q \in I\!PR^4$ are two points in $\Pi$ and $A$ is a symmetric matrix representing a quadratic surface, compute the pole of $p \vee q$ according to the planar section of the surface.

Similarly to the two dimensional case $Ap$ and $Aq$ represent a (spatial) polar of $p$ and $q$ respectively. The difference is that the polar is now a plane instead of a line. However, calculating $(Ap) \wedge (Aq) \wedge \Pi$, leads to the two dimensional pole on $\Pi$ (see section 2.6.2 for all the necessary formulae).

- If $p,q \in I\!PR^4$ are two points in $\Pi$, $A$ is a symmetric matrix representing a quadratic surface and, furthermore, $s \in \Pi$ is a point on the planar intersection of the surface, compute the intersection of the tangent at $s$ and $p \vee q$.

Essentially the same formulae can be used as in (2.52). Indeed, the tangential plane of $A$ is given by $As$ and the intersection of this plane with $\Pi$ will give the tangent in $\Pi$.

As presented later, these formulae (together with the ones listed in 2.6.2) will make it possible to generalise all two dimensional results into 3D. Examples for that will be presented later.