CHAPTER I

SYNTHETIC AND ANALYTIC GEOMETRY

The purpose of this chapter is to review some basic facts from classical deductive geometry and coordinate geometry from slightly more advanced viewpoints. The latter reflect the approaches taken in subsequent chapters of these notes.

1. Axioms for Euclidean geometry

In his book, *Foundations of Geometry*, which was first published in 1900, D. Hilbert (1862–1943) described a set of axioms for classical Euclidean geometry which met modern standards for logical completeness and have been extremely influential ever since. In his formulation, there are six primitive concepts: Points, lines, the notion of one point lying between two others (betweenness), congruence of segments (same distances between the endpoints) and congruence of angles (same angular measurement in degrees or radians). The axioms on these undefined concepts are divided into five classes: Incidence, order, congruence, parallelism and continuity. One notable feature of this classification is that only one class (congruence) requires the use of all six primitive concepts. More precisely, the concepts needed for the axiom classes are given as follows:

Axiom class	**Concepts required**
Incidence	Point, line, plane
Order	Point, line, plane, betweenness
Congruence	All six
Parallelism	Point, line, plane
Continuity	Point, line, plane, betweenness

Strictly speaking, Hilbert's treatment of continuity involves congruence of segments, but the continuity axiom may be formulated without this concept (see Forder, *Foundations of Euclidean Geometry*, p. 297).

As indicated in the table above, congruence of segments and congruence of angles are needed for only one of the axiom classes. Thus it is reasonable to divide the theorems of Euclidean geometry into two classes — those which require the use of congruence and those which do not. Of course, the former class is the more important one in classical Euclidean geometry (it is widely noted that "geometry" literally means "earth measurement"). The main concern of these notes is with theorems of the latter class. Although relatively few theorems of this type were known to the classical Greek geometers and their proofs almost always involved congruence in some way, there is an extensive collection of geometrical theorems having little or nothing to do with congruence.[1]

The viewpoint employed to prove such results contrasts sharply with the usual viewpoint of Euclidean geometry. In the latter subject one generally attempts to prove as much as possible without recourse to the Euclidean Parallel Postulate, and this axiom is introduced only when it is unavoidable. However, in dealing with noncongruence theorems, one assumes the parallel postulate very early in the subject and attempts to prove as much as possible without explicitly discussing congruence. Unfortunately, the statements and proofs of many such theorems are often obscured by the need to treat numerous special cases. **Projective geometry** provides a mathematical framework for stating and proving many such theorems in a simpler and more unified fashion.

2. Coordinate interpretation of primitive concepts

> As long as algebra and geometry proceeded along separate paths, their advance was slow and their applications limited. But when these sciences joined company [through analytic geometry], they drew from each other fresh vitality, and thenceforward marched on at a rapid pace towards perfection. — **J.-L. Lagrange (1736–1813)**

Analytic geometry has yielded powerful methods for dealing with geometric problems. One reason for this is that the primitive concepts of Euclidean geometry have precise numerical formulations in Cartesian coordinates. A *point* in 2- or 3-dimensional coordinate space $\mathbb{R}^2$ or $\mathbb{R}^3$ becomes an ordered pair or triple of real numbers. The *line* joining the points $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$ becomes the set of all $\mathbf{x}$ expressible in vector form as

$$\mathbf{x} \;=\; \mathbf{a} \;+\; t\cdot(\mathbf{b} = \mathbf{a})$$

for some real number t (in $\mathbb{R}^2$ the third coordinate is suppressed). A plane in $\mathbb{R}^3$ is the set of all $\mathbf{x}$ whose coordinates (x_1, x_2, x_3) satisfy a nontrivial linear equation

$$a_1x_1 \;+\; a_2x_2 \;+\; a_3x_3 \;=\; b$$

for three real numbers $a_1,\ a_2\ a_3$ that are not all zero. The point $\mathbf{x}$ is *between* $\mathbf{a}$ and $\mathbf{b}$ if

$$\mathbf{x} \;=\; \mathbf{a} \;+\; t\cdot(\mathbf{b} - \mathbf{a})$$

where the real number t satisfies $0 < t < 1$. Two *segments* are *congruent* if and only if the distances between their endpoints (given by the usual Pythagorean formula) are equal, and two *angles* $\angle\,\mathbf{abc}$ and $\angle\,\mathbf{xyz}$ are *congruent* if their cosines defined by the usual formula

$$\cos\angle\,\mathbf{uvw} \;=\; \frac{(\mathbf{u}-\mathbf{v})\cdot(\mathbf{w}-\mathbf{v})}{|\mathbf{u}-\mathbf{v}|\;|\mathbf{w}-\mathbf{v}|}$$

are equal. We note that the cosine function and its inverse can be defined mathematically without any explicit appeal to geometry by means of the usual power series expansions (for example, see Appendix F in the book by Ryan or pages 182–184 in the book by Rudin; both references are listed in the bibliography).

[1] This discovery did not come from mathematical axiom manipulation for its own sake, but rather from the geometrical theory of *drawing in perspective* begun by Renaissance artists and engineers. See the books by Courant and Robbins, Newman, Kline and Coolidge for more information on the historical origins; some online references are also given in the comments on the 2007 reprinting of these notes, which appear in the Preface.

In the context described above, the axioms for Euclidean geometry reflect crucial algebraic properties of the real number system and the analytic properties of the cosine function and its inverse.

3. Lines and planes in $\mathbb{R}^2$ and $\mathbb{R}^3$

We have seen that the vector space structures on $\mathbb{R}^2$ and $\mathbb{R}^3$ yield convenient formulations for some basic concepts of Euclidean geometry, and in this section we shall see that one can use linear algebra to give a unified description of lines and planes.

THEOREM I.1. *Let $P \subset \mathbb{R}^3$ be a plane, and let $\mathbf{x} \in P$. Then*

$$P(\mathbf{x}) \;=\; \{\ \mathbf{x} \in \mathbb{R}^3 \mid \mathbf{y} \;=\; \mathbf{z} \;-\; \mathbf{x},\ \text{some } \mathbf{z} \in P\ \}$$

is a 2-dimensional vector subspace of $\mathbb{R}^3$. Furthermore, if $\mathbf{v} \in P$ is arbitrary, then $P(\mathbf{v}) = P(\mathbf{x})$.

Proof. Suppose P is defined by the equation $a_1x_1 \;+\; a_2x_2 \;+\; a_3x_3 \;=\; b$. We claim that

$$P(\mathbf{x}) = \{\ \mathbf{y} \in \mathbb{R}^3 \mid a_1y_1 \;+\; a_2y_2 \;+\; a_3y_3 \;=\; 0\ \}\ .$$

Since the coefficients a_i are not all zero, the set $P(\mathbf{x})$ is a 2-dimensional vector subspace of $\mathbb{R}^3$ by Theorem A.10. To prove that $P(\mathbf{x})$ equals the latter set, note that $\mathbf{y} \in P(\mathbf{x})$ implies

$$\sum_{i=1}^{3} a_iy_i \;=\; \sum_{i=1}^{3} a_i(z_i - x_i) \;=\; \sum_{i=1}^{3} a_iz_i \;-\; \sum_{i=1}^{3} a_ix_i \;=\; b - b \;=\; 0$$

and conversely $\sum_{i=1}^{3} a_iy_i \;=\; 0$ implies

$$0 \;=\; \sum_{i=1}^{3} a_iz_i \;-\; \sum_{i=1}^{3} a_ix_i \;=\; \sum_{i=1}^{3} a_iz_i \;-\; b\ .$$

This shows that $P(\mathbf{x})$ is the specified vector subspace of $\mathbb{R}^3$.

To see that $P(\mathbf{v}) = P(\mathbf{x})$, notice that both are equal to $\{\ \mathbf{y} \in \mathbb{R}^3 \mid \sum_{i=1}^{3} a_iy_i \;=\; 0\}$ by the reasoning of the previous paragraph.■

Here is the corresponding result for lines.

THEOREM I.2. *Let $n = 2$ or 3, let $L \subset \mathbb{R}^n$ be a line, and let $\mathbf{x} \in L$. Then*

$$L(\mathbf{x}) \;=\; \{\ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{y} \;=\; \mathbf{z} \;-\; \mathbf{x},\ \text{some } \mathbf{z} \in L\ \}$$

is a 1-dimensional vector subspace of $\mathbb{R}^n$. Furthermore, if $\mathbf{v} \in L$ is arbitrary, then $L(\mathbf{v}) = L(\mathbf{x})$.

Proof. Suppose P is definable as

$$\{\ \mathbf{z} \in \mathbb{R}^n \mid \mathbf{z} \ = \ \mathbf{a} \ - \ t(\mathbf{b} - \mathbf{a}),\ \text{some } t \in \mathbb{R}\ \}$$

where $\mathbf{a} \neq \mathbf{b}$. We claim that

$$L(\mathbf{x}) \ = \ \{\ \mathbf{z} \in \mathbb{R}^{\mathbf{n}} \mid \mathbf{z} \ = \ \mathbf{a} \ - \ \mathbf{t}(\mathbf{b} - \mathbf{a}),\ \text{some } \mathbf{t} \in \mathbb{R}\ \}\ .$$

Since the latter is a 1-dimensional subspace of $\mathbb{R}^n$, this claim implies the first part of the theorem. The second part also follows because both $L(\mathbf{v})$ and $L(\mathbf{x})$ are then equal to this subspace.

Since $\mathbf{x} \in L$, there is a real number s such that $\mathbf{x} \ = \ \mathbf{a} + s(\mathbf{b} - \mathbf{a})$. If $\mathbf{y} \in L(\mathbf{x})$, write $\mathbf{y} = \mathbf{z} - \mathbf{x}$, where $\mathbf{z} \in L$; since $\mathbf{z} \in L$, there is a real number r such that $\mathbf{z} = \mathbf{a} + r(\mathbf{b} - \mathbf{a})$. If we subtract $\mathbf{x}$ from $\mathbf{z}$ we obtain

$$\mathbf{y} \ = \ \mathbf{z} \ - \ \mathbf{x} \ = \ (r - s)(\mathbf{b} \ - \ \mathbf{a})\ .$$

Thus $L(\mathbf{x})$ is contained in the given subspace. Conversely, if $\mathbf{y} = t(\mathbf{b} - \mathbf{a})$, set $\mathbf{z} = \mathbf{x} + \mathbf{y}$. Then

$$\mathbf{z} \ = \ \mathbf{x} \ + \ \mathbf{y} \ = \ \mathbf{a} \ + \ s(\mathbf{b} \ - \ \mathbf{a}) \ + \ t(\mathbf{b} \ - \ \mathbf{a}) \ = \ \mathbf{a} \ + \ (s+t)(\mathbf{b} \ - \ \mathbf{a})\ .$$

Thus $\mathbf{y} \in L(\mathbf{x})$, showing that the given subspace is equal to $L(\mathbf{x})$.■

The following definition will yield a unified reformulation of the theorems above:

Definition. Let V be a vector space over a field $\mathbb{F}$, let $S \subset V$ be a nonempty subset, and let $\mathbf{x} \in V$. The *translate of S by $\mathbf{x}$*, written $\mathbf{x} + S$, is the set

$$\{\ \mathbf{x} \in S \mid \mathbf{y} \ = \ \mathbf{x} \ + \ \mathbf{s},\ \text{some } \mathbf{s} \in S\ \}\ .$$

The fundamental properties of translates are given in the following theorems; the proof of the first is left as an exercise.

THEOREM I.3. *If $\mathbf{z},\ \mathbf{x} \in V$ and $S \subset V$ is nonempty, then $\mathbf{z} \ + \ (\mathbf{x} \ + \ S) \ = \ (\mathbf{z} \ + \ \mathbf{x}) \ + \ S$.*■

THEOREM I.4. *Let V be a vector space, let W be a vector subspace of V, let $\mathbf{x} \in V$, and suppose $\mathbf{y} \in \mathbf{x} \ + \ W$. Then $\mathbf{x} \ + \ W \ = \ \mathbf{y} \ + \ W$.*

Proof. If $\mathbf{z} \in \mathbf{y} \ + \ W$, then $\mathbf{z} = \mathbf{y} + \mathbf{u}$, where $\mathbf{u} \in W$. But $\mathbf{y} = \mathbf{x} + \mathbf{v}$, where $\mathbf{v} \in W$, and hence $\mathbf{z} = \mathbf{x} + \mathbf{u} + \mathbf{v}$, where $\mathbf{u} + \mathbf{v} \in W$. Hence $\mathbf{y} + W \subset \mathbf{x} + W$.

On the other hand, if $\mathbf{z} \in \mathbf{x} + W$, then $\mathbf{z} \in \mathbf{x} + \mathbf{w}$, where $\mathbf{w} \in W$. Since $\mathbf{y} = \mathbf{x} + \mathbf{v}$ (as above), it follows that

$$\mathbf{x} \ + \ \mathbf{w} \ = \ (\mathbf{x} + \mathbf{v}) \ + \ (\mathbf{w} - \mathbf{v}) \ = \ \mathbf{y} \ + \ (\mathbf{w} - \mathbf{v}) \ \in \ \mathbf{y} \ + \ W\ .$$

Consequently, we also have $\mathbf{x} + W \quad \subset \quad \mathbf{y} + W$.■

We shall now reformulate Theorem 1 and Theorem 2.

THEOREM I.5. *Every plane in $\mathbb{R}^3$ is a translate of a 2-dimensional vector subspace, and every line in $\mathbb{R}^2$ or $\mathbb{R}^3$ is a translate of a 1-dimensional vector subspace.*

Proof. If A is a line or plane with $\mathbf{x} \in A$ and $A(\mathbf{x})$ is defined as above, then it is easy to verify that $A \ = \ \mathbf{x} \ + \ A(\mathbf{x})$. ■

The converse to Theorem 5 is also true.

THEOREM I.6. *Every translate of a 2-dimensional subspace of $\mathbb{R}^3$ is a plane, and every translate of a 1-dimensional vector subspace of $\mathbb{R}^2$ or $\mathbb{R}^3$ is a line.*

Proof. CASE 1. Two-dimensional subspaces. Let $\mathbf{b}$ and $\mathbf{c}$ form a basis for W, and let $\mathbf{a} = \mathbf{c} \times \mathbf{b}$ (the cross product; see Section 5 of the Appendix). Then $\mathbf{y} \in W$ if and only if $\mathbf{a} \cdot \mathbf{y} = 0$ by Theorem A.10 and the cross product identities at the beginning of Section 5 of the Appendix). We claim that $\mathbf{z} \in \mathbf{x} \in W$ if and only if $\mathbf{a} \cdot \mathbf{z} = \mathbf{a} \cdot \mathbf{x}$.

If $\mathbf{z} \in \mathbf{x} + W$, write $\mathbf{z} = \mathbf{x} + \mathbf{w}$, where $\mathbf{w} \in W$. By distributivity of the dot product we have

$$\mathbf{a} \cdot \mathbf{z} \;=\; \mathbf{a} \cdot (\mathbf{x} + \mathbf{w}) \;=\; (\mathbf{a} \cdot \mathbf{x}) \;+\; (\mathbf{a} \cdot \mathbf{w}) \;=\; (\mathbf{a} \cdot \mathbf{x})$$

the latter following because $\mathbf{a} \cdot \mathbf{w} = 0$. Conversely, if $\mathbf{a} \cdot \mathbf{z} = (\mathbf{a} \cdot \mathbf{x})$, then

$$\mathbf{a} \cdot (\mathbf{w} - \mathbf{x}) \;=\; (\mathbf{a} \cdot \mathbf{z}) \;-\; (\mathbf{a} \cdot \mathbf{x}) \;=\; 0$$

and hence $\mathbf{z} - \mathbf{x} \in W$. Since $\mathbf{z} = \mathbf{x} + (\mathbf{z} \cdot \mathbf{x})$, clearly $\mathbf{x} \in \mathbf{z} + W$.

CASE 2. One-dimensional subspaces. Let $\mathbf{w}$ be a nonzero (hence spanning) vector in W, and let $\mathbf{y} \in \mathbf{x} + \mathbf{w}$. Then the line $\mathbf{xy}$ is equal to $\mathbf{x} + W$.■

The theorems above readily yield an alternate characterization of lines in $\mathbb{R}^2$ which is similar to the characterization of planes in $\mathbb{R}^3$.

THEOREM I.7. *A subset of $\mathbb{R}^2$ is a line if ane only if there exist $a_1,\ a_2,\ b \in \mathbb{R}$ such that not both a_1 and a_2 are zero and the point $\mathbf{x} = (x_1, x_2)$ lies in the subset if and only if $a_1x_1 + a_2x_2 = b$.*

Proof. Suppose that the set S is defined by the equation above. Let W be the set of all $\mathbf{y} = (y_1, y_2)$ such that $a_1yx_1 + a_2y_2 = 0$. By Theorem A.10 we know that W is a 1-dimensional subspace of $\mathbb{R}^2$. Thus if $\mathbf{y} \in W$, the argument proving Theorem 1 shows that $S = \mathbf{y} + W$.

On the other hand, suppose that $\mathbf{y} + W$ is a line in $\mathbb{R}^2$, where W is a 1-dimensional vector subspace of $\mathbb{R}^2$. Let $\mathbf{w} = (w_1, w_2)$ be a nonzero vector in W; then $J(\mathbf{w}) = (w_2, -w_1)$ is also nonzero, and $\mathbf{z} \in W$ if and only if it is perpendicular to $J(\mathbf{w})$ by Theorem A.10. A modified versionof the proof of Theorem 6, Case 1, shows that $\mathbf{x} \in \mathbf{y} + W$ if and only if

$$J(\mathbf{w}) \cdot \mathbf{x} \;=\; J(\mathbf{w}) \cdot \mathbf{y}\ .$$

Thus it suffices to take $(a_1, a_2) = J(\mathbf{w})$ and $b = \mathbf{w} \cdot \mathbf{y}$. ■

EXERCISES

1. Prove Theorem 3.

2. Verify the assertion $S \;=\; \mathbf{x} + S(\mathbf{x})$ made in Theorem 5.

3. Let V be a vector space, let $W \subset V$ be a vector subspace, and suppose that $\mathbf{u}$ and $\mathbf{v}$ are vectors in V. Prove that the sets $\mathbf{u} + W$ and $\mathbf{v} + W$ are either disjoint or equal.

4. Fill in the details of the proof of Theorem 7.

5. Let P be the unique plane through the given triples of points in each of the following cases. Find an equation defining P, and determine the 2-dimensional vector subspace of which P is a translate.

(i) $(1,3,2)$, $(4,1,-1)$, $(2,0,0)$.

(ii) $(1,1,0)$, $(1,0,1)$, $(0,1,1)$.

(iii) $(2,-1,3)$, $(1,1,1)$, $(3,0,4)$.

6. Suppose that we are given two distinct lines L, M in $\mathbb{R}^3$ which meet at the point $\mathbf{x}_0$, and write these lies as $L = \mathbf{x}_0 + V$ and $M = \mathbf{x}_0 + W$, were V and W have bases given by $\{\,\mathbf{a}\,\}$ and $\{\,\mathbf{b}\,\}$ respectively. Explain why there is a plane containint L and M [*Hint:* Why do $\mathbf{a}$ and $\mathbf{b}$ span a 2-dimensional vector subspace?]

7. Let $\{\,\mathbf{a},\ \mathbf{b},\ \mathbf{c}\,\}$ be a basis for $\mathbb{R}^3$, let V be the 1-dimensional vector subspace spanned by $\mathbf{a}$, and let W be the 1-dimensional vector space spanned by $\mathbf{c}-\mathbf{b}$. Prove that the lines V and $\mathbf{b}+W$ have no points in common. [*Hint:* If such a point exists, then by the preceding exercise the two lines are coplanar and lie in some plane $\mathbf{z}+X$, where X is a 2-dimensional vector subspace. Why do $\mathbf{0}$, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ all lie in X, and why does this imply that $\mathbf{z}+X = X$? Derive a contradiction from this and the preceding two sentences.]

CHAPTER II

AFFINE GEOMETRY

In the previous chapter we indicated how several basic ideas from geometry have natural interpretations in terms of vector spaces and linear algebra. This chapter continues the process of formulating basic geometric concepts in such terms. It begins with standard material, moves on to consider topics not covered in most courses on classical deductive geometry or analytic geometry, and it concludes by giving an abstract formulation of the concept of geometrical incidence and closely related issues.

1. Synthetic affine geometry

In this section we shall consider some properties of Euclidean spaces which only depend upon the axioms of incidence and parallelism

Definition. A *three-dimensional incidence space* is a triple $(S, \mathcal{L}, \mathcal{P})$ consisting of a nonempty set S (whose elements are called *points*) and two nonempty disjoint families of proper subsets of S denoted by $\mathcal{L}$ (*lines*) and $\mathcal{P}$ (*planes*) respectively, which satisfy the following conditions:

(I – 1) Every line (element of $\mathcal{L}$) contains at least two points, and every plane (element of $\mathcal{P}$) contains at least three points.

(I – 2) If $\mathbf{x}$ and $\mathbf{y}$ are distinct points of S, then there is a unique line L such that $\mathbf{x},\ \mathbf{y} \in L$.

Notation. The line given by **(I –2)** is called $\mathbf{xy}$.

(I – 3) If $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are distinct points of S and $\mathbf{z} \notin \mathbf{xy}$, then there is a unique plane P such that $\mathbf{x},\ \mathbf{y},\ \mathbf{z} \in P$.

(I – 4) If a plane P contains the distinct points $\mathbf{x}$ and $\mathbf{y}$, then it also contains the line $\mathbf{xy}$.

(I – 5) If P and Q are planes with a nonempty intersection, then $P \cap Q$ contains at least two points.

Of course, the standard example in $\mathbb{R}^3$ with lines and planes defined by the formulas in Chapter I (we shall verify a more general statement later in this chapter). A list of other simple examples appears in Prenowitz and Jordan, *Basic Concepts of Geometry*, pp. 141–146.

A few theorems in Euclidean geometry are true for every three-dimensional incidence space. The proofs of these results provide an easy introduction to the synthetic techniques of these notes. In the first six results, the triple $(S, \mathcal{L}, \mathcal{P})$ denotes a fixed three-dimensional incidence space.

Definition. A set B of points in S is *collinear* if there is some line L in S such that $B \subset L$, and it is *noncollinear* otherwise. A set A of points in S is *coplanar* if there is some plane P in S such that $A \subset P$, and it is *noncoplanar* otherwise. — Frequently we say that the points $\mathbf{x},\ \mathbf{y},\ \cdots$ (*etc.*) are collinear or coplanar if the set with these elements is collinear or coplanar respectively.

THEOREM II.1. *Let* $\mathbf{x}$*,* $\mathbf{y}$ *and* $\mathbf{z}$ *be distinct points of* S *such that* $\mathbf{z} \notin \mathbf{xy}$*. Then* $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ *is a noncollinear set.*

Proof. Suppose that L is a line containing the given three points. Since $\mathbf{x}$ and $\mathbf{y}$ are distinct, by $(\mathbf{I}-\mathbf{2})$ we know that $L = \mathbf{xy}$. By our assumption on L it follows that $\mathbf{z} \in L$; however, this contradicts the hypothesis $\mathbf{z} \notin \mathbf{xy}$. Therefore there is no line containing $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$.■

THEOREM II.2. *There is a subset of four noncoplanar points in* S*.*

Proof. Let P be a plane in S. We claim that P contains three noncollinear points. By $(\mathbf{I}-\mathbf{1})$ we know that P contains three distinct points $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}_0$. If these three points are noncollinear, let $\mathbf{c} = \mathbf{c}_0$. If they are collinear, then the line L containing them is a subset of P by $(\mathbf{I}-\mathbf{4})$, and since $\mathcal{L}$ and $\mathcal{P}$ are disjoint it follows that L must be a proper subset of P; therefore there is some point $\mathbf{c} \in P$ such that $\mathbf{c} \notin L$, and by the preceding result the set $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is noncollinear. Thus in any case we know that P contains three noncollinear points.

Since P is a proper subset of S, there is a point $\mathbf{d} \in S$ such that $\mathbf{d} \notin P$. We claim that $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\ \mathbf{d}\}$ is noncoplanar. For if Q were a plane containing all four points, then $\mathbf{a}, \mathbf{b}, \mathbf{c} \in P$ would imply $P = Q$, which contradicts our basic stipulation that $\mathbf{d} \notin P$.■

THEOREM II.3. *The intersection of two distinct lines in* S *is either a point or the empty set.*

Proof. Suppose that $\mathbf{x} \neq \mathbf{y}$ but both belong to $L \cap M$ for some lines L and M. By property $(\mathbf{I}-\mathbf{2})$ we must have $L = M$. Thus the intersection of distinct lines must consist of at most one point.■

THEOREM II.4. *The intersection of two distinct planes in* S *is either a line or the empty set.*

Proof. Suppose that P and Q are distinct planes in S with a nonempty intersection, and let $\mathbf{x} \in P \cap Q$. By $(\mathbf{I}-\mathbf{5})$ there is a second point $\mathbf{y} \in P \cap Q$. If L is the line $\mathbf{xy}$, then $L \subset P$ and $L \subset Q$ by two applications of $(\mathbf{I}-\mathbf{4})$; hence we have $L \subset P \cap Q$. If there is a point $\mathbf{z} \in P \cap Q$ with $\mathbf{z} \notin L$, then the points $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are noncollinear but contained in both of the planes P and Q. By $(\mathbf{I}-\mathbf{3})$ we must have $P = Q$. On the other hand, by assumption we know $P \neq Q$, so we have reached a contradiction. The source of this contradiction is our hypothesis that $P \cap Q$ strictly contains L, and therefore it follows that $P \cap Q = L$.■

THEOREM II.5. *Let* L *and* M *be distinct lines, and assume that* $L \cap M \neq \varnothing$*. Then there is a unique plane* P *such that* $L \subset P$ *and* $M \subset P$*.*

In less formal terms, given two intersecting lines there is a unique plane containing them.

Proof. Let $\mathbf{x} \in L \cap M$ be the unique common point (it is unique by Theorem 3). By $(\mathbf{I}-\mathbf{2})$ there exist points $\mathbf{y} \in L$ and $\mathbf{z} \in M$, each of which is distinct from $\mathbf{x}$. The points $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are noncollinear because $L = \mathbf{xy}$ and $\mathbf{z} \in M - \{\mathbf{x}\} = M - L$. By $(\mathbf{I}-\mathbf{3})$ there is a unique plane P such that $\mathbf{x}, \mathbf{y}, \mathbf{z} \in P$, and by $(\mathbf{I}-\mathbf{4})$ we know that $L \subset P$ and $M \subset P$. This proves the existence of a plane containing both L and M. To see this plane is unique, observe that every plane Q containing both lines must contain $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. By $(\mathbf{I}-\mathbf{3})$ there is a unique such plane, and therefore we must have $Q = P$.■

THEOREM II.6. *Given a line L and a point $\mathbf{z}$ not on L, there is a unique plane P such that $L \subset P$ and $\mathbf{z} \in P$.*

Proof. Let $\mathbf{x}$ and $\mathbf{y}$ be distinct points of L, so that $L = \mathbf{xy}$. We then know that the set $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is noncollinear, and hence there is a unique plane P containing them. By **(I − 4)** we know that $L \subset P$ and $\mathbf{z} \in P$. Conversely, if Q is an arbitrary plane containing L and $\mathbf{z}$, then Q contains the three noncollinear points $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, and hence by **(I − 3)** we know that $Q = P$.■

Notation. We shall denote the unique plane in the preceding result by $L\mathbf{z}$.

Of course, all the theorems above are quite simple; their conclusions are probably very clear intuitively, and their proofs are fairly straightforward arguments. One must add Hilbert's Axioms of Order or the Euclidean Parallelism Axiom to obtain something more substantial. Since our aim is to introduce the parallel postulate at an early point, we might as well do so now (a thorough treatment of geometric theorems derivable from the Axioms of Incidence and Order appears in Chapter 12 of Coxeter, *Introduction to Geometry*; we shall discuss the Axioms of Order in Section VI.6 of these notes).

Definition. Two lines in a three-dimensional incidence space S are *parallel* if they are disjoint and coplanar (note particularly the second condition). If L and L' are parallel, we shall write $L||L'$ and denote their common plane by LL'.

Note that if $L||M$ then $M||L$ because the conditions in the definition of parallelism are symmetric in the two lines.

Affine three-dimensional incidence spaces

Definition. A three-dimensional incidence space $(S, \mathcal{L}, \mathcal{P})$ is an *affine three-space* if the following holds:

(EPP) *For each line L in S and each point $\mathbf{x} \notin L$ there is a unique line $L' \subset L\mathbf{x}$ such that $\mathbf{x} \in L$ and $L \cap L' = \varnothing$ (in other words, there is a unique line L' which contains $\mathbf{x}$ and is parallel to L).*

This property is often called *the Euclidean Parallelism Property*, the *Euclidean Parallel Postulate* or *Playfair's Postulate*, and its significance for the axiomatic foundations of geometry was recognized by Proclus Diadochus (412–485); see the previously cited online reference for background on these individuals.

A discussion of the origin of the term "affine" appears in Section II.5 of the following online site:

`http://math.ucr.edu/~res/math133/geometrynotes2b.pdf`

Many nontrivial results in Euclidean geometry can be proved for arbitrary affine three-spaces. We shall limit ourselves to two examples here and leave others as exercises. In Theorems 7 and 8 below, the triple $(S, \mathcal{L}, \mathcal{P})$ will denote an arbitrary affine three-dimensional incidence space.

THEOREM II.7. *Two lines which are parallel to a third line are parallel.*

Proof. There are two cases, depending on whether or not all three lines lie in a single plane; to see that the three lines need not be coplanar in ordinary 3-dimensional coordinate geometry, consider the three lines in $\mathbb{R}^3$ given by the z-axis and the lines joining $(1,0,0)$ and $(0,1,0)$ to $(1,0,1)$ and $(0,1,1)$ respectively.

THE COPLANAR CASE. Suppose that we have three distinct lines L, M, N in a plane P such that $L||N$ and $M||N$; we want to show that $L||N$.

If L is not parallel to N, then there is some point $\mathbf{x} \in L \cap N$, and it follows that L and N are distinct lines through $\mathbf{x}$, each of which is parallel to M. However, this contradicts the Euclidean Parallel Postulate. Therefore the lines L and N cannot have any points in common.

THE NONCOPLANAR CASE. Let α be the plane containing L and M, and let β be the plane containing M and N. By the basic assumption in this case we have $\alpha \neq \beta$. We need to show that $L \cap N = \varnothing$ but L and N are coplanar.

The lines L and N are disjoint. Assume that the L and N have a common point that we shall call $\mathbf{x}$. Let γ be the plane determined by $\mathbf{x}$ and N (since $L||M$ and $\mathbf{x} \in L$, clearly $\mathbf{x} \notin M$). Since $\mathbf{x} \in L \subset \alpha$ and $M \subset \alpha$, Theorem 6 implies that $\alpha = \gamma$. A similar argument shows that $\beta = \gamma$ and hence $\alpha = \beta$; the latter contradicts our basic stipulation that $\alpha \neq \beta$, and therefore it follows that L and N cannot have any points in common.

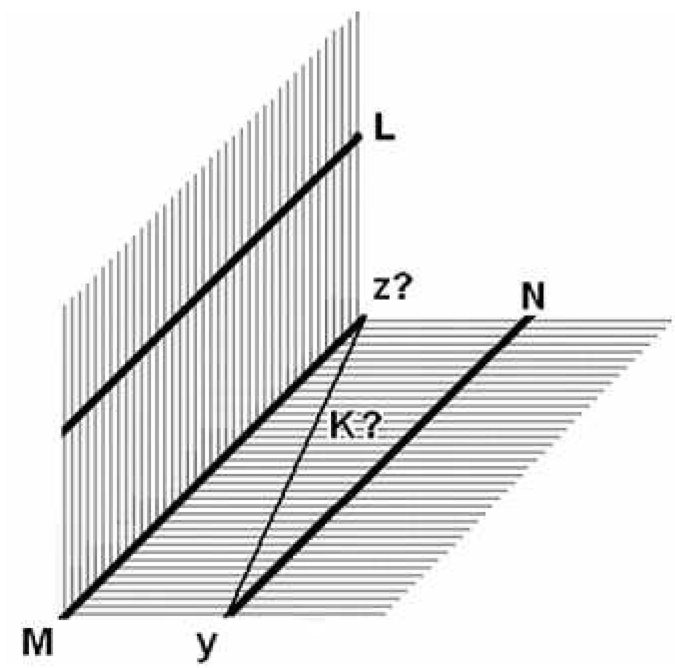


Figure II.1

The lines L and N are coplanar. Let $\mathbf{y} \in N$, and consider the plane $L\mathbf{y}$. Now L cannot be contained in β because $\beta \neq \alpha = LM$ and $M \subset \beta$. By construction the planes $L\mathbf{y}$ and β have the point $\mathbf{y}$ in common, and therefore we know that $L\mathbf{y}$ meets β in some line K. Since L and K are coplanar, it will suffice to show that $N = K$. Since N and K both contain $\mathbf{y}$ and all three lines M, N and K are contained in β, it will suffice to show that $K||M$.

Suppose the lines are not parallel, and let $\mathbf{z} \in K \cap M$. Since $L||M$ it follows that $\mathbf{z} \notin L$. Furthermore, $L \cup K \subset L\mathbf{y}$ implies that $\mathbf{z} \in L\mathbf{y}$, and hence $\mathbf{y} = L\mathbf{z}$. Since $\mathbf{z} \in M$ and L and M

are coplanar, it follows that $M \subset L\mathbf{z}$. Thus M is contained in $L\mathbf{y} \cap \beta$, and since the latter is the line K, this shows that $M = K$. On the other hand, by construction we know that $M \cap N = \varnothing$ and $K \cap N \neq \varnothing$, so that M and K are obviously distinct. This contradiction implies that $K || M$ must hold.■

The next result is an analog of the Parallel Postulate for parallel planes.

THEOREM II.8. *If P is a plane and $\mathbf{x} \notin P$, then there is a unique plane Q such that $\mathbf{x} \in Q$ and $P \cap Q = \varnothing$.*

Proof. Let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c} \in P$ be the noncollinear points, and consider the lines A', B' through $\mathbf{x}$ which are parallel to $A = \mathbf{bc}$ and $B = \mathbf{ac}$. Let Q be the plane determined by A' and B', so that $\mathbf{x} \in Q$ by hypothesis. We claim that $P \cap Q = \varnothing$.

Assume the contrary; since $\mathbf{x} \in QA$ and $\mathbf{x} \notin P$, the intersection $P \cap Q$ is a line we shall call L.

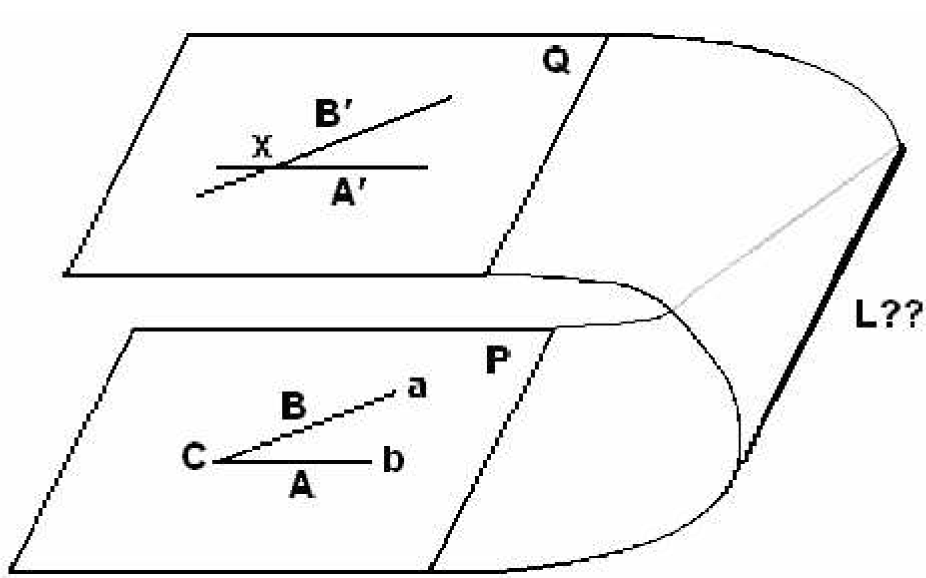


Figure II.2

Step 1. We shall show that $L \neq A$, B. The proof that $L \neq A$ and $L \neq B$ are similar, so we shall only show $L \neq A$ and leave the other part as an exercise. — If $L = A$, then $L \subset Q$. Since A' is the unique line in Q which is parallel to A, there must be a point $\mathbf{u} \in B' \cap A$. Consider the plane $B'\mathbf{c}$. Since $\mathbf{c} \in A$, it follows that $A \subset B'\mathbf{c}$. Hence $B'\mathbf{c}$ is a plane containing A and B. The only plane which satisfies these conditions is P, and hence $B' \subset P$. But $\mathbf{x} \in B'$ and $\mathbf{x} \notin P$, so we have a contradiction. Therefore we must have $L \neq A$.

Step 2. We claim that either $A' \cap L$ and $A \cap L$ are both nonempty or else $B' \cap L$ and $B \cap L$ are both nonempty. — We shall only show that if either $A' \cap L$ is $A \cap L$ is empty then both $B' \cap L$ and $B \cap L$ are both nonempty, since the other case follows by reversing the roles of A and B. Since L and A both lie in the plane P, the condition $A \cap L = \varnothing$ implies $A||L$. Since $A||A'$, by Theorem 7 and Theorem 1 we know that $A'||L$. Since $B \neq A$ is a line through the point $\mathbf{c} \in A$, either $B = L$ or $B \cap L \neq \varnothing$ holds by the Parallel Postulate (in fact, $B \neq L$ by Step

1). Likewise, B and B' are lines through $\mathbf{x}$ in the plane Q and $L \subset Q$, so that the $A'||L$ and the Parallel Postulate imply $B' \cap L \neq \varnothing$.

Step 3. There are two cases, depending upon whether $A' \cap L$ and $A \cap L$ are both nonempty or $B' \cap L$ and $B \cap L$ are both nonempty. Only the latter will be considered, since the former follows by a similar argument. Let $\mathbf{y} \in B \cap L$ and $\mathbf{z} \in B' \cap L$; since $B \cap B' = \varnothing$, if follows that $\mathbf{y} \neq \mathbf{z}$ and hence $L = \mathbf{yz}$. Let β be the plane BB'. Then $L \subset \beta$ since $\mathbf{z},\ \mathbf{y} \in \beta$. Since $L \neq B$, the plane β is the one determined by L and B. But $L,\ B \subset P$ by assumption, and hence $\beta = P$. In other words, B' is contained in P. But $\mathbf{x} \in B'$ and $\mathbf{x} \notin P$, a contradiction which shows that the line L cannot exist.■

Following standard terminology, we shall say that the plane Q is *parallel* to P or that it is the plane parallel to P which passes through $\mathbf{x}$.

Corresponding definitions for incidence planes and affine planes exist, and analogs of Theorems 1, 2, 3 and 7 hold for these objects. However, incidence planes have far fewer interesting properties than their three-dimensional counterparts, and affine planes are best studied using the methods of projective geometry that are developed in later sections of these notes.

EXERCISES

Definition. A line and a plane in a three-dimensional incidence space are *parallel* if they are disjoint.

Exercises **1–4** *are to be proved for arbitrary* 3*-dimensional incidence spaces.*

1. Suppose that each of two intersecting lines is parallel to a third line. Prove that the three lines are coplanar.

2. Suppose that the lines L and L' are coplanar, and there is a line M not in this plane such that $L||M$ and $L'||M$. Prove that $L||L'$.

3. Let P and Q be planes, and assume that each line in P is parallel to a line in Q. Prove that P is parallel to Q.

4. Suppose that the line L is contained in the plane P, and suppose that $L||L'$. Prove that $L'||P$ or $L \subset P$.

In exercises **5–6**, *assume the incidence space is affine.*

5. Let P and Q be parallel planes, and let L be any line which contains a point of Q and is parallel to a line in P. Prove that L is contained in Q. [*Hint:* Let M be the line in P, and let $\mathbf{x} \in L \cap Q$. Prove that $L = M\mathbf{x} \cap Q$.]

6. Two lines are said to be **skew lines** if they are not coplanar. Suppose that L and M are skew lines. Prove that there is a unique plane P such that $L \subset P$ and P is parallel to M. [*Hint:* Let $\mathbf{x} \in L$, let M' be a line parallel to M which contains $\mathbf{x}$, and consider the plane LM'.]

2. Affine subspaces of vector spaces

Let $\mathbb{F}$ be a field, and let V be a vector space over $\mathbb{F}$ (in fact, everything in this section goes through if we take $\mathbb{F}$ to be a **skew-field** as described in Section 1 of Appendix A). Motivated by Section I.3, we define *lines* and *planes* in V to be translates of 1- and 2-dimensional vector subspaces of V. Denote these families of lines and planes by $\mathcal{L}_V$ and $\mathcal{P}_V$ respectively. If $\dim V \geq 3$ we shall prove that $(V, \mathcal{L}_V, \mathcal{P}_V)$ satisfies all the conditions in the definition of an affine incidence 3-space except perhaps for the incidence axiom $\mathbf{I-5}$, and we shall show that the latter also holds if $\dim V = 3$.

Theorem II.9. *If V, etc. are as above and $\dim V \geq 3$, then $\mathcal{L}_V$ and $\mathcal{P}_V$ are nonempty disjoint families of proper subsets of V.*

Proof. Since $\dim V \geq 3$ there are 1- and 2-dimensional vector subspaces of V, and therefore the families $\mathcal{L}_V$ and $\mathcal{P}_V$ are both nonempty. If we have $Y \in \mathcal{L}_V \cap \mathcal{P}_V$, then we may write

$$Y \; = \; \mathbf{x} \; + \; W_1 \; = \; \mathbf{y} \; + \; W_2$$

where $\dim W_i = i$. By Theorem I.4 we know that $\mathbf{y} \in \mathbf{x}+W_1$ implies the identity $\mathbf{x}+W_1 = \mathbf{y}+W_1$, and therefore Theorem I.3 implies

$$W_2 \; = \; -\mathbf{y} \; + \; (\mathbf{y} \; + \; W_1) \; = \; -\mathbf{y} \; + \; (\mathbf{y} \; + \; W_2) \; = \; W_2 \; .$$

Since $\dim W_1 \neq \dim W_2$ this is impossible. Therefore the families $\mathcal{L}_V$ and $\mathcal{P}_V$ must be disjoint. To see that an element of either family is a proper subset of V, suppose to the contrary that $\mathbf{x} + W = V$, where $\dim W = 1$ or 2. Since $\dim W < 3 \leq \dim V$, it follows that W is a proper subset of V; let $\mathbf{v} \in V$ be such that $\mathbf{v} \notin W$. By our hypothesis, we must have $\mathbf{x} + \mathbf{v} \in \mathbf{x} + W$, and thus we also have

$$\mathbf{v} \; = \; -\mathbf{x} \; + \; (\mathbf{x} \; + \; \mathbf{v}) \; \in \; -\mathbf{x} \; + \; (\mathbf{x} \; + \; W) \; = \; W$$

which contradicts our fundamental condition on $\mathbf{x}$. The contradiction arises from our assumption that $\mathbf{x}+W = V$, and therefore this must be false; therefore the sets in $\mathcal{L}_V$ and $\mathcal{P}_V$ are all proper subsets of V.■

Theorem II.10. *Every line in V contains at least two points, and every plane contains at least three points.*

Proof. Let $\mathbf{x}+W$ be a line or plane in V, and let $\{\,\mathbf{w}_1\,\}$ or $\{\,\mathbf{w}_1,\ \mathbf{w}_2\,\}$ be a basis for W depending upon whether $\dim W$ equals 1 or 2. Take the subsets $\{\,\mathbf{v},\ \mathbf{v}+\mathbf{w}_1,\,\}$ or $\{\,\mathbf{v},\ \mathbf{v}+\mathbf{w}_1,\ \mathbf{v}+\mathbf{w}_2\,\}$ in these respective cases.■

Theorem II.11. *Given two distinct points in V, there is a unique line containing them.*

Proof. Let $\mathbf{x} \neq \mathbf{y}$ be distinct points in V, and let L_0 be the 1-dimensional vector subspace spanned by the nonzero vector $\mathbf{y} - \mathbf{x}$. Then $L = \mathbf{x} + L_0$ is a line containing $\mathbf{x}$ and $\mathbf{y}$. Suppose now that M is an arbitrary line containing $\mathbf{x}$ and $\mathbf{y}$. Write $M = \mathbf{z} + W$ where $\dim W = 1$. Then Theorem I.4 and $\mathbf{x} \in M = \mathbf{z} + W$ imply that $M = \mathbf{x} + W$. Furthermore, $\mathbf{y} \in M = \mathbf{x} + W$ then implies that $\mathbf{y} - \mathbf{x} \in W$, and since the latter vector spans L_0 it follows that $L_0 \subset W$. However, $\dim L_0 = \dim W$, and therefore $L_0 = W$ (see Theorem A.8). Thus the line $M = \mathbf{z} + W$ must be equal to $\mathbf{x} + L_0 = L$.■

Theorem II.12. *Given three points in V that are not collinear, there is a unique plane containing them.*

Proof. Let $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ be the noncollinear points. If $\mathbf{y}-\mathbf{x}$ and $\mathbf{z}-\mathbf{x}$ were linearly dependent, then there would be a 1-dimensional vector subspace W containing them and hence the original three points would all lie on the line $\mathbf{x}+W$. Therefore we know that $\mathbf{y}-\mathbf{x}$ and $\mathbf{z}-\mathbf{x}$ are **linearly independent**, and thus the vector subspace W they span is 2-dimensional. If $P = \mathbf{x}+W$, then it follows immediately that P is a plane containing $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. To prove uniqueness, suppose that $\mathbf{v}+U$ is an arbitrary plane containing all three points. As before, we must have $\mathbf{v}+U = \mathbf{x}+U$ since $bfx \in \mathbf{v}+U$, and since we also have $\mathbf{y},\ \mathbf{z} \in \mathbf{v}+U = \mathbf{x}+U$ it also follows as in earlier arguments that $\mathbf{y}-\mathbf{x}$ and $\mathbf{z}-\mathbf{x}$ lie in U. Once again, since the two vectors in question span the subspace W, it follows that $W \subset U$, and since the dimensions are equal it follows that $W = U$. Thus we have $\mathbf{v}+U = \mathbf{x}+W$, and hence there is only one plane containing the original three points.■

Theorem II.13. *If P is a plane in V and $\mathbf{x},\ \mathbf{y} \in P$, then the unique line containing $\mathbf{x}$ and $\mathbf{y}$ is a subset of P.*

Proof. As before we may write $P = \mathbf{x}+W$, where W is a 2-dimensional subspace; we also know that the unique line joining $\mathbf{x}$ and $\mathbf{y}$ has the form $L = \mathbf{x}+L_0$, where L_0 is spanned by $\mathbf{y}-\mathbf{x}$. The condition $\mathbf{y} \in P$ implies that $\mathbf{y}-\mathbf{x} \in W$, and since W is a vector subspace it follows that $L_0 \subset W$. But this immediately implies that $L = \mathbf{x}+L_0 \subset \mathbf{x}+W = P$, which is what we wanted to prove.■

Theorem II.14. (Euclidean Parallelism Property) *Let L be a line in V, and let $\mathbf{y} \notin L$. Then there is a unique line M such that (i) $\mathbf{y} \in M$, (ii) $L \cap M = \varnothing$, (iii) L and M are coplanar.*

Proof. Write $L = \mathbf{x}+L_0$ where $\dim L_0 = 1$, and consider the line $M = \mathbf{y}+L_0$. Then M clearly satisfies the first condition. To see it satisfies the second condition, suppose to the contrary that there is some common point $\mathbf{z} \in L \cap M$. Then the identities

$$\mathbf{z} \ \in \ L \ = \ \mathbf{x} \ + \ L_0 \qquad \mathbf{z} \ \in \ M \ = \ \mathbf{y} \ + \ L_0$$

imply that $L = \mathbf{x}+L_0 = \mathbf{z}+L_0 = \mathbf{y}+L_0 = M$, which contradicts the basic conditions that $\mathbf{y} \in M$ but $\mathbf{y} \notin L$. Therefore $L \cap M = \varnothing$. Finally, to see that M also satisfies the third condition, let W be the subspace spanned by L_0 and $\mathbf{y}-\mathbf{x}$; the latter vector does not belong to L_0 because $\mathbf{y} \notin L$, and therefore W must be a 2-dimensional vector subspace of V. If we now take $P = \mathbf{x}+W$, it follows immediately that $L \subset P$ and also

$$M \ = \ \mathbf{y} \ + \ L_0\, \mathbf{x} \ + \ (\mathbf{y}-\mathbf{x}) \ + \ L_0 \ \subset \ \mathbf{x} \ + \ W \ = \ P$$

so that L and M are coplanar. Therefore M satisfies all of the three conditions in the theorem.

To complete the proof, we need to show that *there is only one line which satisfies all three conditions in the theorem.* — In any case, we know there is only one plane which contains L and $\mathbf{y}$, and hence it must be the plane $P = \mathbf{x}+W$ from the preceding paragraph. Thus if N is a line satisfying all three conditions in the theorem, it must be contained in P. Suppose then that $N = \mathbf{y}+L_1$ is an arbitrary line in P with the required properties. Since $\mathbf{y}+L_1 \in \mathbf{x}+W = \mathbf{y}+W$, it follows that $L_1 \subset W$. Therefore, if $\mathbf{0} \neq \mathbf{u} \in L_1$ we can write $\mathbf{u} = s(\mathbf{y}-\mathbf{x})+t\mathbf{v}$, where $\mathbf{v}$ is a nonzero vector in L_0 and $s,\ t \in \mathbb{F}$. The assumption that $L \cap N = \varnothing$ implies there are no scalars p and q such that

$$\mathbf{x} \ + \ p\mathbf{v} \ = \ \mathbf{y} \ + \ q\mathbf{u}$$

holds. Substituting for $\mathbf{u}$ in this equation, we may rewrite it in the form

$$p\mathbf{v} \ = \ (\mathbf{y} - \mathbf{x}) \ + \ q\mathbf{u} \ = \ (1 + sq)(\mathbf{y} - \mathbf{x}) \ + \ qt\mathbf{v}$$

and hence by equating coefficients we cannot find p and q such that $p = qt$ and $1 + sq = 0$. Now if $s \neq 0$, then these two equations have the solution $q = -s^{-1}$ and $p = -s^{-1}t$. Therefore, if there is no solution then we must have $s = 0$. The latter in turn implies that $\mathbf{u} = t\mathbf{v}$ and hence $L_1 = L_0$, so that $N = \mathbf{y} + L_1 = \mathbf{y} + L_0 = M$.■

THEOREM II.15. *If* $\dim V = 3$ *and two planes in* V *have a nonempty intersection, then their intersection contains a line.*

Proof. Let P and Q be planes, and let $\mathbf{x} \in P \cap Q$. Write $P = \mathbf{x} + W$ and $Q = \mathbf{x} + U$, where $\dim W = \dim U = 2$. Then

$$P \cap Q \ = \ (\mathbf{x} + W) \ \cap \ (\mathbf{x} + U)$$

clearly contains $\mathbf{x} + (W \cap U)$, so it suffices to show that the vector subspace $W \cap U$ contains a 1-dimensional vector subspace. However, we have

$$\dim(W \cap U) \ = \ \dim W \ - \ \dim U \ - \ \dim(W + U) \ = \ 4 \ - \ \dim(W + U)$$

and since $\dim(W + U) \leq \dim V = 3$, the displayed equation immediately implies $\dim(W \cap U) \geq 4 - 3 = 1$. Hence the intersection of the vector subspaces is nonzero, and as such it contains a nonzero vector $\mathbf{z}$ as well as the 1-dimensional subspace spanned by $\mathbf{z}$.■

The preceding results imply that Theorems 7 and 8 from Section 1, and the conclusions of Exercises 5 and 6 from that section, are all true for the models $(V, \mathcal{L}_V, \mathcal{P}_V)$ described above provided $\dim V = 3$. In some contexts it is useful to interpret the conclusions of the theorems or exercises in temrs of the vector space structure on V. For example, in Theorem 8 if P is the plane $\mathbf{x} + W$, then the parallel plane Q will be $\mathbf{y} + W$. Another example is discussed in Exercise 5 below.

Generalizing incidence to higher dimensions

The characterization of lines and planes as translates of 1- and 2-dimensional subspaces suggests a simple method for generalizing incidence structures to dimensions greater than three.[1] Namely, define a *k-plane* in a vector space V to be a translate of a k-dimensional vector subspace.

The following quotation from Winger, *Introduction to Projective Geometry*,[2] may help the reader understand the reasons for formulating the concepts of affine geometry in arbitrary dimensions.[3]

[1]The explicit mathematical study of higher dimensional geometry began around the middle of the 19^{th} century, particularly in the work of L. Schläfli (1814–1895). Many ideas in his work were independently discovered by others with the development of linear algebra during the second half of that century.

[2]See page 15 of that book.

[3]Actually, spaces of higher dimensions play an important role in theoretical physics. Einstein's use of a four-dimensional space-time is of course well-known, but the use of spaces with dimensions ≥ 4 in physics was at least implicit during much of the 19^{th} century. In particular, 6-dimensional configuration spaces were implicit in work on celestial mechanics, and spaces of assorted other dimensions were widely used in classical dynamics.

> The timid may console themselves with the reflection that the geometry of four and higher dimensions is, if not a necessity, certainly a convenience of language — a translation of the algebra — and let the philosophers ponder the metaphysical questions involved in the idea of a point set of higher dimensions.

We shall conclude this section with a characterization of k-planes in V, where V is finite-dimensional and $1+1 \neq 0$ in $\mathbb{F}$; in particular, the result below applies to the $V = \mathbb{R}^n$. An extension of this characterization to all fields except the field $\mathbb{Z}_2$ with two elements is given in Exercise 1 below.

Definition. Let V be a vector space over the field $\mathbb{F}$, and let $P \subset V$. We shall say P is a *flat subset* of V if for each pair of distinct points $\mathbf{x},\ \mathbf{y} \in \mathbb{F}$ the line $\mathbf{xy}$ is contained in $\mathbb{F}$.

THEOREM II.16. *Let V be a finite-dimensional vector space over a field $\mathbb{F}$ in which $1+1 \neq 0$. Then a nonempty set $P \subset V$ is a flat subset if and only if it is a k-plane for some integer k satisfying $0 \leq k \leq \dim V$.*

Definition. A subset $S \subset V$ is said to be an *affine subspace* if can be written as $\mathbf{x}+W$, where $\mathbf{x} \in V$ and W is a vector subspace of V. With this terminology, we can restate the theorem to say that if $1+1 \neq 0$ in $\mathbb{F}$, then a nonempty subset of V is an affine subspace if and only if it is a flat subset.

Proof. We split the proof into the "if" and "only if" parts.

Every k-plane is a flat subset. Suppose that W is a k-dimensional vector subspace and $\mathbf{x} \in V$. Let $\mathbf{y},\ \mathbf{z} \in \mathbf{x}+V$. Then we may write $\mathbf{y} = \mathbf{x}+\mathbf{u}$ and $\mathbf{z} = \mathbf{x}+\mathbf{v}$ for some distinct vectors $\mathbf{u},\ \mathbf{v} \in W$. A typical point on the line $\mathbf{yz}$ has the form $\mathbf{y}+t(\mathbf{z}-\mathbf{y})$ for some scalar t, and we have

$$\mathbf{y} \ + \ t(\mathbf{z}-\mathbf{y}) \ = \ \mathbf{x} \ + \ \mathbf{u} \ + \ t(\mathbf{v}-\mathbf{u}) \ \in \ \mathbf{x} \ + \ W$$

which shows that $\mathbf{x}+W$ is a flat subset. Note that this implication does not require any assumption about the nontriviality of $1+1$.

Every flat subset has the form $\mathbf{x}+W$ for some vector subspace W. If we know this, we also know that $k = \dim W$ is less than or equal to $\dim V$. — Suppose that $\mathbf{x} \in P$, and let $W = (-\mathbf{x})+P$; we need to show that W is a vector subspace. To see that W is closed under scalar multiplication, note first that $\mathbf{w} \in W$ implies $\mathbf{x}+\mathbf{w} \in P$, so that flatness implies every point on the line $\mathbf{x}(\mathbf{x}+\mathbf{w})$ is contained in P. For each scalar t we know that $\mathbf{x}+t\mathbf{w}$ lies on this line, and thus each point of this type lies in $P = \mathbf{x}+W$. If we subtract $\mathbf{x}$ we see that $t\mathbf{w} \in W$ and hence W is closed under scalar multiplication.

To see that W is closed under vector addition, suppose that $\mathbf{w}_1$ and $\mathbf{w}_2$ are in W. By the previous paragraph we know that $2\mathbf{w}_1$ and $2\mathbf{w}_2$ also belong to W, so that $\mathbf{u}_1 = \mathbf{x}+2\mathbf{w}_1$ and $\mathbf{u}_2 = \mathbf{x}+2\mathbf{w}_2$ are in P. Our hypothesis on $\mathbb{F}$ implies the latter contains an element $\frac{1}{2} = (1+1)^{-1}$, so by flatness we also know that

$$\tfrac{1}{2}\,(\,\mathbf{u}_1+\mathbf{u}_2\,) \ + \ \tfrac{1}{2}\,(\,\mathbf{u}_2-\mathbf{u}_1\,) \ \in \ P \ .$$

If we now expand and simplify the displayed vector, we see that it is equal to $\mathbf{v}+\mathbf{w}_1+\mathbf{w}_2$. Therefore it follows that $\mathbf{w}_1+\mathbf{w}_2 \in W$, and hence W is a vector subspace of V.■

Not surprisingly, if S is an affine subspace of the finite-dimensional vector space V, then we define its *dimension* by $\dim S = \dim W$, where W is a vector subspace of V such that $S = \mathbf{x}+W$.

— *This number is well defined* because $S = \mathbf{x} + W = \mathbf{y} + U$ implies $\mathbf{y} \in \mathbf{x} + W$, so that $\mathbf{y} + W = \mathbf{x} + W = \mathbf{y} + U$ and hence

$$W \;=\; -\mathbf{y} \;+\; (\mathbf{y} + W) \;=\; -\mathbf{y} \;+\; (\mathbf{y} + U) \;=\; U \;.\blacksquare$$

Hyperplanes

One particularly important family of affine subspaces in a finite-dimensional vector space V is the set of all **hyperplanes** in V. We shall conclude this section by defining such objects and proving a crucial fact about them.

Definition. Let n be a positive integer, and let V be an n-dimensional vector space over the field $\mathbb{F}$. A subset $H \subset V$ is called a *hyperplane* in V if H is the translate of an $(n-1)$-dimensional subspace. — In particular, if $\dim V = 3$ then a hyperplane is just a plane and if $\dim V = 2$ then a hyperplane is just a line.

One reason for the importance of hyperplanes is that if $k < \dim V$ then every k-plane is an intersection of finitely many hyperplanes (see Exercise II.5.4).

We have seen that planes in $\mathbb{R}^3$ and lines in $\mathbb{R}^2$ are describable as the sets of all points $\mathbf{x}$ which satisfy a nontrivial first degree equation in the coordinates (x_1, x_2, x_3) or (x_1, x_2) respectively (see Theorems I.1, I.5 and I.7). The final result of this section is a generalization of these facts to arbitrary hyperplanes in $\mathbb{F}^n$, where $\mathbb{F}$ is an arbitrary field.

THEOREM II.17. *In the notation of the previous paragraph, let H be a nonempty subset of $\mathbb{F}^n$. Then H is a hyperplane if and only if there exist scalars $c_1, \cdots, c_n$ which are not all zero such that H is the set of all $\mathbf{x} \in \mathbb{F}^n$ whose coordinates $x_1, \cdots, x_n$ satisfy the equation*

$$\sum_{i=0}^{n} c_i\, x_i \;=\; b$$

for some $b \in \mathbb{F}$.

Proof. Suppose that H is defined by the equation above. If we choose j such that $a_j \neq 0$ and let $\mathbf{e}_j$ be the unit vector whose j^{th} coordinate is 1 and whose other coordinates are zero, then we have $-a_j^{-1} b\mathbf{e}_j \in H$, and hence the latter is nonempty. Set W equal to $(-\mathbf{z}) + H$, where $\mathbf{z} \in H$ is fixed. As in Chapter I, it follows that $\mathbf{y} \in W$ if and only if its coordinates $y_1, \cdots, y_n$ satisfy the equation $\sum_i a_i y_i = 0$. Since the coefficients a_i are not all zero, it follows from Theorem A.10 that W is an $(n-1)$-dimensional vector subspace of $\mathbb{F}^n$, and therefore $H = \mathbf{x} + W$ is a hyperplane.$\square$

Conversely, suppose H is a hyperplane and write $\mathbf{x} + W$ for a suitable vector $\mathbf{x}$ and $(n-1)$-dimensional subspace W. Let $\{\mathbf{w}_1, \cdots, \mathbf{w}_{n-1}\}$ be a basis for W, and write these vectors out in coordinate form:

$$\mathbf{w}_i \;=\; \big(w_{i,1}, \;\cdots\;, w_{i,n}\big)$$

If B is the matrix whose rows are the vectors $\mathbf{w}_i$, then the rank of B is equal to $(n-1)$ by construction. Therefore, by Theorem A.10 the set Y of all $\mathbf{y} = (y_1, \cdots, y_n)$ which solve the system

$$\sum_j y_j w_{i,j} \;=\; 0 \qquad\qquad \big(1 \leq i \leq (n-1)\big)$$

is a 1-dimensional vector subspace of $\mathbb{F}^n$. Let $\mathbf{a}$ be a nonzero (hence spanning) vector in Y.

We claim that $\mathbf{z} \in W$ if and only if $\sum_i a_i z_i = 0$. By construction, W is contained in the subspace S of vectors whose coordinates satisfy this equation. By Theorem A.10 we know that $\dim S = (n-1)$, which is equal to $\dim W$ by our choice of the latter; therefore Theorem A.9 implies that $W = S$, and it follows immediately that H is the set of all $\mathbf{z} \in \mathbb{F}^n$ whose coordinates $z_1, \cdots, z_n$ satisfy the nontrivial linear equation $\sum_i a_i z_i = \sum_i a_i x_i$ (where $\mathbf{x} = (x_1, \cdots, x_n) \in W$ is the fixed vector chosen as in the preceding paragraph).■

EXERCISES

1. Prove that Theorem 16 remains true for every field **except** $\mathbb{Z}_2$. Give an example of a flat subspace of $(\mathbb{Z}_2)^3$ which is not a k-plane for some k.

2. Let $\mathbb{F}$ be a field. Prove that the lines in $\mathbb{F}^2$ defined by the equations $ax + by + cz = 0$ and $a'x + by + c'z = 0$ (compare Theorem I.7 and Theorem 17 above) are parallel or identical if and only if $ab' - ba' = 0$.

3. Find the equation of the hyperplane in $\mathbb{R}^3$ passing through the (noncoplanar) points $(1,0,1,0)$, $(0,1,0,1)$, $(0,1,1,0)$, and $(1,0,0,1)$.

4. Suppose that $\mathbf{x}_1 + W_1$ and $\mathbf{x}_2 + W_2$ are k_1- and k_2-planes in a vector space V such that $\mathbf{x}_1 + W_1 \cap \mathbf{x}_2 + W_2 \neq \varnothing$. Let $\mathbf{z}$ be a common point of these subsets. Prove that their intersection is equal to $\mathbf{z} + W_1 \cap \mathbf{z} + W_2$, and generalize this result to arbitrary finite intersections.

5. Let V be a 3-dimensional vector space over $\mathbb{F}$, and suppose we are given the configuration of Exercise II.1.6 (L and M are skew lines, and P is a plane containing L but parallel to M). Suppose that the skew lines are given by $\mathbf{x} + W$ and $\mathbf{y} + U$. Prove that the plane P is equal to $\mathbf{x} + (U + W)$ [*Hint:* Show that the latter contains L and is disjoint from M.].

6. Suppose that $\dim V = n$ and $H = \mathbf{x} + W$ is a hyperplane in V. Suppose that $\mathbf{y} \in V$ but $\mathbf{y} \notin H$. Prove that $H' = \mathbf{y} + W$ is the unique hyperplane K such that $\mathbf{y} \in K$ and $H \cap K = \varnothing$. [*Hints:* If $\mathbf{z} \in H \cap H'$ then $N = \mathbf{x} + W = \mathbf{z} + W = \mathbf{y} + W = H'$. If $K = \mathbf{y} + U$ where U is some $(n-1)$-dimensional vector subspace different from W, explain why $\dim W \cap U = n - 2$. Choose a basis A of $(n-2)$ vectors for this intersection, and let $\mathbf{u}_0 \in U$, $\mathbf{w}_0 \in W$ such that $\mathbf{u}_0, \mathbf{w}_0 \notin U \cap W$. Show that $A \cup \{\mathbf{u}_0, \mathbf{w}_0\}$ is a basis for V, write $\mathbf{y} - \mathbf{x}$ in terms of this basis, and use this equation to find a vector which lies in $H \cap K = (\mathbf{x} + W) \cap (\mathbf{y} + U)$.]

3. Affine bases

We shall need analogs of linear independence and spanning that apply to arbitrary k-planes and not just k-dimensional vector subspaces. The starting points are two basic observations.

THEOREM II.18. *Suppose that S is a k-plane in a vector space V over the field $\mathbb{F}$. Given $\mathbf{a}_1, \cdots, \mathbf{a}_r \in S$, let $t_1, \cdots, t_r \in S$ be such that $\sum_j t_j = 1$. Then $\sum_j t_j\mathbf{a}_j \in S$.*

Proof. Write $S = \mathbf{x} + W$, where $\mathbf{x} \in S$ and W is a k-dimensional vector subspace, and for each i write $\mathbf{a}_i = \mathbf{x} + \mathbf{w}_i$ where $\mathbf{w}_i \in \mathbb{F}$. Then

$$\sum_j t_j\mathbf{a}_j = \sum_j t_j(\mathbf{x} + \mathbf{w}_j) = \sum_j t_j\mathbf{x} + \sum_j t_j\mathbf{w}_j = \mathbf{x} + \sum_j t_j\mathbf{w}_j$$

where the latter holds because $\sum_j t_j = 1$. Since W is a vector subspace, we know that $\sum_j t_j\mathbf{w}_j \in W$, and therefore it follows that $\sum_j t_j\mathbf{a}_j \in W$.■

THEOREM II.19. *If V is as above and $T \subset V$ is an arbitrary subset, define the* **affine hull** *of T by*

$$\mathcal{H}(T) = \{ \mathbf{x} \in V \mid \mathbf{x} = \sum_j t_j\mathbf{w}_j, \text{ where } \mathbf{v}_i \in T \text{ for all } t \text{ and } \sum_j t_j = 1\}$$

(note that the sum is finite, but T need not be finite). Then $\mathcal{H}(T)$ is an affine subspace of V.

Sometimes we shall also say that $\mathcal{H}(T)$ is the *affine span* of T, and we shall say that T *affinely spans* an affine subspace S if $S = \mathcal{H}(T)$.

Proof. Suppose that $\mathbf{x},\ \mathbf{y} \in \mathcal{H}(T)$, and write

$$\mathbf{x} = \sum_i s_i\mathbf{u}_i = \sum_j t_j\mathbf{v}_j$$

where $\mathbf{u}_i \in T$ and $\mathbf{v}_j \in T$ for all i and j, and the coefficients satisfy $\sum_i s_i = \sum_j t_j = 1$. We need to show that $\mathbf{x} + c(\mathbf{y} - \mathbf{x}) \in \mathcal{H}(T)$. But

$$\mathbf{x} + c(\mathbf{y} - \mathbf{x}) = \sum_i s_i\mathbf{u}_i + c \cdot \left(\sum_j t_j\mathbf{v}_j - \sum_i s_i\mathbf{u}_i \right) =$$

$$(1-c) \cdot \sum_i s_i\mathbf{u}_i + c \cdot \sum_j t_j\mathbf{v}_j = \sum_i s_i\,(1-c)\mathbf{u}_i + \sum_j t_j\,c\mathbf{v}_j\ .$$

We have no *a priori* way of knowing whether any of the vectors $\mathbf{u}_i$ are equal to any of the vectors $\mathbf{v}_j$, but in any case we can combine like terms to rewrite the last expression as $\sum_q r_q\mathbf{w}_q$, where the vectors $\mathbf{w}_q$ run through all the vectors $\mathbf{u}_i$ and $\mathbf{v}_j$, and the coefficients r_q are given accordingly; by construction, we then have

$$\sum_q r_q = \sum_i s_i\,(1-c) + \sum_j t_j\,c = (1-c) \cdot \sum_i s_i + c \cdot \sum_j t_j = (1-c) \cdot 1 + c \cdot 1 = 1$$

and therefore it follows that the point $\mathbf{x} + c(\mathbf{y} - \mathbf{x})$ belongs to $\mathcal{H}(T)$.■

Thus *a linear combination of points in an affine subspace will also lie in the subspace* **provided** *the coefficients add up to* 1, and by Theorem 19 this is the most general type of linear combination one can expect to lie in S.

Definition. A vector $\mathbf{v}$ is an *affine combination* of the vectors $\mathbf{x}_0, \cdots, \mathbf{x}_n$ if we have $\mathbf{x} = \sum_j t_j\mathbf{x}_j$, where $\sum_j t_j = 1$. Thus the affine hull $\mathcal{H}(T)$ of a set T is the set of all (finite) affine combinations of vectors in T.

AFFINE VERSUS LINEAR COMBINATIONS. If a vector $\mathbf{y}$ is a linear combination of the vectors $\mathbf{v}_1 < \cdots, \mathbf{v}_n$, then it is automatically an affine combination of $\mathbf{0}, \mathbf{v}_1, \cdots, \mathbf{v}_n$, for if $\mathbf{y} = \sum_i t_i\mathbf{x}_i$ then

$$\mathbf{y} = \left(1 - \sum_j t_j\right) \cdot \mathbf{0} + \sum_i t_i\mathbf{x}_i .$$

Definition. Let V be a vector space over a field $\mathbb{F}$. and let $X \subset V$ be the set $\{\mathbf{x}_0, \cdots, \mathbf{x}_n\}$. We shall say that X is *affinely dependent* if one element of X is expressible as an affine combination of the others and *affinely independent* otherwise. By convention, one point subsets are affinely independent.

The next result gives the fundamental relationship between affine dependence and linear dependence (and, by taking negations, it also gives the fundamental relationship between affine independence and linear independence).

THEOREM II.20. *In the setting above, the finite set $X = \{\mathbf{x}_0, \cdots, \mathbf{x}_n\} \subset V$ is affinely dependent if and only if the set $X' = \{\mathbf{x}_1 - \mathbf{x}_0, \cdots, \mathbf{x}_n - \mathbf{x}_0\}$ is linearly dependent. Likewise, the finite set $X = \{\mathbf{x}_0, \cdots, \mathbf{x}_n\} \subset V$ is affinely independent if and only if the set $X' = \{\mathbf{x}_1 - \mathbf{x}_0, \cdots, \mathbf{x}_n - \mathbf{x}_0\}$ is linearly independent.*

Proof. Since affine dependence and affine independence are the negations of each other and similarly for linear dependence and linear independence, it will suffice to prove the first conclusion in the theorem.

Proof that X is affinely dependent if X' is linearly dependent. By linear dependence there is some $k > 0$ such that

$$\mathbf{x}_k - \mathbf{x}_0 = \sum_{i \neq 0,k} c_i\,(\mathbf{x}_i - \mathbf{x}_0)$$

and therefore we also have

$$\mathbf{x}_k = \mathbf{x}_0 + \sum_{i \neq 0,k} c_i\,\mathbf{x}_i - \sum_{i \neq 0,k} c_i\,\mathbf{x}_0 = \left(1 - \sum_{i \neq 0,k} c_i\right)\mathbf{x}_0 + \sum_{i \neq 0,k} c_i\,\mathbf{x}_i .$$

Therefore $\mathbf{x}_k$ is also an affine combination of all the $\mathbf{x}_j$ such that $j \neq k$.

Proof that X' is linearly dependent if X is affinely dependent. By assumption there is some k such that

$$\mathbf{x}_k = \sum_{j \neq k} c_j\,\mathbf{x}_j$$

where $\sum_{j \neq k} \; c_j = 1$. Therefore we also have

$$\mathbf{x}_k - \mathbf{x}_0 \;=\; \left(\sum_{j \neq k} c_j \, \mathbf{x}_j \right) \;-\; \mathbf{x}_0 \;=\; \sum_{j \neq k} c_j (\mathbf{x}_j - \mathbf{x}_0) \ .$$

Note that we can take the summation on the right hand side to run over all j such that $j \neq k, \ 0$ because $\mathbf{x}_0 - \mathbf{x}_0 = \mathbf{0}$.

There are now two cases depending on whether $k > 0$ or $k = 0$. In the first case, we have obtained an expression for $\mathbf{x}_k - \mathbf{x}_0$ in terms of the other vectors in X', and therefore X' is linearly dependent. Suppose now that $k = 0$, so that the preceding equation reduces to

$$\mathbf{0} \;=\; \sum_{j > 0} c_j (\mathbf{x}_j - \mathbf{x}_0) \ .$$

Since $\sum_{j > 0} \; c_j \;\; = \;\; 1$ it follows that $c_m \neq 0$ for some $m > 0$, and this now implies

$$\mathbf{x}_m - \mathbf{x}_0 \;=\; \sum_{j \neq m, 0} \frac{-c_j}{c_m} (\mathbf{x}_j - \mathbf{x}_0)$$

which shows that X' is linearly dependent.■

One important characterization of linear independence for a set Y is that an arbitrary vector has at most one expression as a linear combination of the vectors in Y. There is a similar characterization of affine independence.

THEOREM II.21. *A (finite) set X of vectors in a given vector space V is affinely independent if and only if every vector in V has* **at most one** *expression as an affine combination of vectors in X.*

Proof. Suppose X is affinely independent and that

$$\mathbf{y} \;=\; \sum_j t_j \mathbf{v}_j \;=\; \sum_j s_j \mathbf{v}_j$$

where $\mathbf{v}_j$ runs through the vectors in V and

$$\sum_j t_j \;=\; \sum_j s_j \;=\; 1 \ .$$

Then we have

$$\left(\sum_j t_j \mathbf{v}_j \right) \;-\; \mathbf{v}_0 \;=\; \left(\sum_j s_j \mathbf{v}_j \right) \;-\; \mathbf{v}_0$$

which in turn implies

$$\left(\sum_j t_j \mathbf{v}_j \right) \;-\; \left(\sum_j t_j \right) \mathbf{v}_0 \;=\; \left(\sum_j s_j \mathbf{v}_j \right) \;-\; \left(\sum_j s_j \right) \mathbf{v}_0$$

so that we have

$$\sum_{j > 0} t_j (\mathbf{v}_j - \mathbf{v}_0) \;=\; \sum_j s_j (\mathbf{v}_j - \mathbf{v}_0) \ .$$

Since the vectors $\mathbf{v}_j - \mathbf{v}_0$ (where $j > 0$) is linearly independent, it follows that $t_j = s_j$ for all $j > 0$. Once we know this, we can also conclude that $t_0 = 1 - \sum_{j>0} t_j$ is equal to $1 - \sum_{j>0} s_j = s_0$, and therefore all the corresponding coefficients in the two expressions are equal.□

Conversely, suppose X satisfies the uniqueness condition for affine combinations, and suppose that we have

$$\sum_{j>0} c_j(\mathbf{v}_j - \mathbf{v}_0) \;=\; \mathbf{0}\;.$$

We then need to show that $c_j = 0$ for all j. But the equation above implies that

$$\mathbf{v}_0 \;=\; \left(\sum_{j>0} c_j(\mathbf{v}_j - \mathbf{v}_0)\right) \;+\; \mathbf{v}_0$$

and if we simplify the right hand side we obtain the equation

$$\mathbf{v}_0 \;=\; \left(1 \;-\; \sum_{j>0} c_j\right)\mathbf{v}_0 \;+\; \sum_{j>0} c_j\mathbf{v}_j\;.$$

The coefficients on both sides add up to 1, so by the uniqueness assumption we must have $c_j = 0$ for all $j > 0$; but this implies that the vectors $\mathbf{v}_j - \mathbf{v}_0$ (where $j > 0$) are linearly independent.■

Definition. If S is an affine subspace of the finite-dimensional vector space V and $T \subset S$ is a finite subset, then T is said to be an *affine basis* for S if T is affinely independent and affinely spans T.

There is a fundamental analog of the preceding results which relates affine bases of affine subspaces and vector space bases of vector subspaces.

THEOREM II.22. *Let V be a finite-dimensional vector space, let S be an affine subspace, and suppose that $S = \mathbf{z} + W$ for a suitable vector $\mathbf{z}$ and vector subspace V. Then the finite set $X = \{\mathbf{x}_0, \cdots, \mathbf{x}_m\} \subset S$ is an affine basis for S if and only if the set $X' = \{\mathbf{x}_1 - \mathbf{x}_0, \cdots, \mathbf{x}_n - \mathbf{x}_0\}$ is linear basis for W.*

Proof. First of all, since $\mathbf{x}_0 \in S$ we may write $S = \mathbf{x}_0 + W$ and forget about the vector $\mathbf{z}$.

Suppose that $X = \{\mathbf{x}_0, \cdots, \mathbf{x}_m\}$ is an affine basis for S, and let $\mathbf{y} \in W$. Since $\mathbf{x}_0 + \mathbf{y} \in \mathbf{x}_0 + W = S$, there exist $s_0, \cdots, s_m \in \mathbb{F}$ such that $\sum_i s_i = 1$ and $\mathbf{x}_0 + \mathbf{y} = \sum_i s_i\mathbf{x}_i$. Subtracting $\mathbf{x}_0$ from both sides and using the equation $\sum_i s_i = 1$, we see that

$$\mathbf{y} \;=\; \sum_{i>0} s_i(\mathbf{x}_i - \mathbf{x}_0)$$

and hence X' spans W. Since X' is linearly independent by Theorem 20, it follows that X' is a basis for W.□

Conversely, suppose that $X' = \{\mathbf{x}_1 - \mathbf{x}_0, \cdots, \mathbf{x}_n - \mathbf{x}_0\}$ is linear basis for W. Since X' is linearly independent, by Theorem 20 we also know that X is affinely independent. To see that X affinely spans S, let $\mathbf{u} \in S$, and write $\mathbf{u} = \mathbf{x}_0 + \mathbf{v}$, where $\mathbf{v} \in S$. Since X' spans W we know that

$$\mathbf{u} \;=\; \mathbf{x}_0 \;+\; \sum_{i>0} s_i(\mathbf{x}_i - \mathbf{x}_0)$$

for appropriate scalars s_i, and if we set $s_0 = 1 - \sum_{i>0} s_i$, then we may rewrite the right hand side of the preceding equation as $\sum_{i\geq 0} s_i\mathbf{x}_i$, where by construction we have $\sum_{i\geq 0} s_i = 1$. Therefore X affinely spans S, and it follows that X must be an affine basis for X,■

Definition. Suppose that we are in the setting of the theorem and X is an affine basis for S, so that each $\mathbf{y} \in S$ can be uniquely written as an affine combination $\sum_i t_i\mathbf{x}_i$, where $\sum_i t_i = 1$. Then the unique coefficients t_i are called the *barycentric coordinates* of $\mathbf{y}$ with respect to X. The physical motivation for this name is simple: Suppose that $\mathbb{F}$ is the real numbers and we place weights $w_i > 0$ at the vectors $\mathbf{v}_i$ in S such that the total weight is w units. Let $t_i = w_i/w$ be the normalized weight at $\mathbf{v}_i$; then $\sum_i t_i\mathbf{x}_i$ is the center of gravity for the resulting physical system (a version of this is true even if one allows some of the coefficients t_i to be negative).

In analogy for linear bases of vector subspaces, the number of elements in an affine basis for an affine subspace S depends on S itself. However, as illustrated by the final result of this section, there is a crucial difference in the formula relating the dimension of S to the number of elements in an affine basis.

THEOREM II.23. *If V is a finite-dimensional vector space and S is an affine subspace, then S has an affine basis. Furthermore, if we write $S = \mathbf{y} + W$ for suitable $\mathbf{y}$ and W, then every affine basis for S has exactly* $\dim W + 1$ *elements.*

Proof. If $X = \{\mathbf{x}_0, \cdots, \mathbf{x}_m\} \subset S$ is an affine basis for $S = \mathbf{y} + W = \mathbf{x}_0 + W$, then $X' = \{\mathbf{x}_1 - \mathbf{x}_0, \cdots, \mathbf{x}_n - \mathbf{x}_0\}$ is a linear basis for W by Theorem 22, and conversely. Therefore the existence of an affine basis for S follows from Theorem 22 and the existence of a linear basis for W. Furthermore, since every linear basis for W contains exactly $\dim W$ elements, by Theorem 22 we know that every affine basis for S contains exactly $\dim W + 1$ elements.■

EXERCISES

In the exercises below, assume that all vectors lie in a fixed finite-dimensional vector space V over a field $\mathbb{F}$.

1. Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$ (a vector space over some field) be noncollinear and for $i = 1, 2, 3$ let $\mathbf{x}_i = t_i\mathbf{a} + u_i\mathbf{b} + v_i\mathbf{c}$, where $t_i + u_i + v_i = 1$. Prove that the points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are collinear if and only if

$$\begin{vmatrix} t_1 & u_1 & v_1 \\ t_2 & u_2 & v_2 \\ t_3 & u_3 & v_3 \end{vmatrix} = 0$$

where the left hand side is a 3×3 determinant.

2. Prove the **Theorem of Menelaus**:[4] Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$ (a vector space over some field) be noncollinear, and suppose we have points $\mathbf{p} \in \mathbf{bc}$, $\mathbf{q} \in \mathbf{ac}$, and $\mathbf{r} \in \mathbf{ab}$. Write these three vectors as

$$\begin{aligned} \mathbf{p} &= \mathbf{b} + t(\mathbf{c} - \mathbf{b}) = t\mathbf{c} + (1-t)\mathbf{b} \\ \mathbf{q} &= \mathbf{a} + u(\mathbf{c} - \mathbf{a}) = u\mathbf{c} + (1-u)\mathbf{a} \\ \mathbf{r} &= \mathbf{a} + v(\mathbf{b} - \mathbf{a}) = v\mathbf{b} + (1-v)\mathbf{a} \end{aligned}$$

[4]MENELAUS OF ALEXANDRIA (*c.* 70 A. D–*c.* 130 A. D.) worked in geometry and astronomy, and he is particularly given credit for making numerous contributions to spherical geometry and trigonometry.

where $t,\ u,\ v$ are appropriate scalars. Then $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ are collinear if and only if

$$\frac{t}{1-t}\cdot\frac{u}{1-u}\cdot\frac{v}{1-v} \;=\; -1\ .$$

3. Prove the **Theorem of Ceva:**[5] In the setting of the preceding exercise, the lines $\mathbf{ap}$, $\mathbf{bq}$ and $\mathbf{cr}$ are concurrent (there is a point which lies on all three lines) if and only if

$$\frac{t}{1-t}\cdot\frac{u}{1-u}\cdot\frac{v}{1-v} \;=\; +1\ .$$

4. Let $\mathbf{a},\ \mathbf{b},\ \mathbf{c} \in V$ (as above) be noncollinear, and suppose we have points $\mathbf{y} \in \mathbf{ba}$ and $\mathbf{x} \in \mathbf{bc}$ which are distinct from $\mathbf{a},\ \mathbf{b},\ \mathbf{c}$ and satisfy these three vectors as

$$\begin{aligned} \mathbf{x} \;&=\; \mathbf{b} \;+\; t(\mathbf{a}-\mathbf{b}) \;=\; t\mathbf{a} \;+\; (1-t)\mathbf{b} \\ \mathbf{y} \;&=\; \mathbf{b} \;+\; u(\mathbf{c}-\mathbf{a}) \;=\; u\mathbf{c} \;+\; (1-u)\mathbf{b} \end{aligned}$$

where $t,\ u$ are appropriate scalars (neither of which is 0 or 1). Prove that the lines $\mathbf{ay}$ and $\mathbf{cx}$ have a point in common if and only if $ut \neq 1$. [*Hint:* Explain why the lines have no points in common if and only if $\mathbf{y}-\mathbf{a}$ and $\mathbf{x}-\mathbf{c}$ are linearly dependent. Write both of these vectors as linear combinations of $\mathbf{a}-\mathbf{b}$ and $\mathbf{c}-\mathbf{b}$, and show that if $\mathbf{z}$ and $\mathbf{w}$ are linearly independent, then $p\mathbf{z}+q\mathbf{w}$ and $r\mathbf{z}+s\mathbf{w}$ are linearly dependent if and only if $sp = rq$. Finally, compare the two conclusions in the preceding sentence.]

5. Let V be a finite-dimensional vector space over the field $\mathbb{F}$, let $W \subset V$ be a vector subspace, suppose that $\dim W = k$, and let $X = \{\mathbf{x}_1,\ \cdots\ ,\mathbf{x}_m\}$ be a finite subset of W. Prove that W is a basis for W in the sense of linear algebra if and only if $X \cup \{\mathbf{0}\}$ is an affine basis for $W \;=\; \mathbf{0}+W$ if the latter is viewed as a k-plane.

6. In the setting of Exercises 2 and 3, suppose that the lines $\mathbf{ap}$, $\mathbf{bq}$ and $\mathbf{cr}$ are concurrent with $t = \frac{1}{2}$ and $v = 1-u$. Express the common point $\mathbf{g}$ of these three lines as a linear combination of $\mathbf{a}$ and $\mathbf{d}$ with the coefficients expressed in terms of u.

---

[5] Giovanni Ceva (1647–1734) is known for the result bearing his name, his rediscovery of Menelaus' Theorem, and his work on hydraulics.

4. Affine bases

In this section we shall generalize certain classical theorems of Euclidean geometry to affine planes over fields in which $1 + 1 \neq 0$. Similar material is often presented in many mathematics courses as *proofs of geometric theorems using vectors.* In fact, the uses of vectors in geometry go far beyond yielding alternate proofs for some basic results in classical Greek geometry; they are often the method of choice for studying all sorts of geometrical problems ranging from purely theoretical questions to carrying out the computations needed to create high quality computer graphics. We shall illustrate the uses of vector algebra in geometry further by proving some nonclassical theorems that figure significantly in the next two chapters.

Let $\mathbb{F}$ be a field in which $1 + 1 \neq 0$, and set $\frac{1}{2}$ equal to $(1 + 1)^{-1}$. Given a vector space V over $\mathbb{F}$ and two distinct vectors $\mathbf{a},\ \mathbf{b} \in V$, the *midpoint* of $\mathbf{a}$ and $\mathbf{b}$ is the vector $\frac{1}{2}(\mathbf{a} + \mathbf{b})$.

THEOREM II.24. *Let V and $\mathbb{F}$ as above, and let $\mathbf{a},\ \mathbf{b},\ \mathbf{c}$ be noncollinear points in V. Then the lines joining $\mathbf{x} = \mathsf{midpoint}(\mathbf{a}, \mathbf{b})$ and $\mathbf{y} = \mathsf{midpoint}(\mathbf{a}, \mathbf{c})$ is parallel to $\mathbf{bc}$.*

In ordinary Euclidean geometry one also knows that the length of the segment joining $\mathbf{x}$ to $\mathbf{y}$ is half the length of the segment joining $\mathbf{b}$ to $\mathbf{c}$ (the length of the segment is just the distance between the endpoints). We do not include such a conclusion because our setting does not include a method for defining distances (in particular, an arbitrary field has no *a priori* notion of distance).

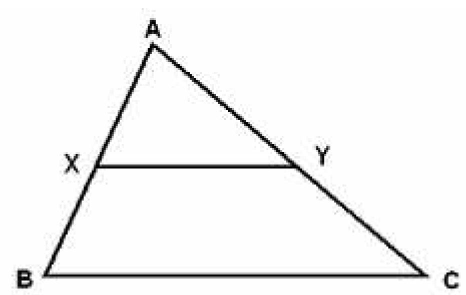


Figure II.3

Proof, Let W be the subspace spanned by $\mathbf{c} - \mathbf{b}$; it follows that $\mathbf{bc} = \mathbf{b} + W$. On the other hand, the line joining the midpoints is given by $\frac{1}{2}(\mathbf{b} + \mathbf{c}) + U$, where U is the subspace spanned by

$$\tfrac{1}{2}(\mathbf{a} + \mathbf{c}) \ - \ \tfrac{1}{2}(\mathbf{a} + \mathbf{b}) \ = \ \tfrac{1}{2}(\mathbf{c} - \mathbf{b}) \ .$$

Since there is a vector $\mathbf{w}$ such that W is spanned by $\mathbf{w}$ and U is spanned by $\frac{1}{2}\mathbf{w}$, clearly $W = U$, and therefore the line joining the midpoints is parallel to $\mathbf{bc}$ by the construction employed to prove Theorem 14.■

Definition. Let V and $\mathbb{F}$ as above, and let $\mathbf{a},\ \mathbf{b},\ \mathbf{c}$ be noncollinear points in V. The *affine triangle* $\Delta\,\mathbf{abc}$ is given by $\mathbf{ab}\ \cup\ \mathbf{ac}\ \cup\ \mathbf{bc}$, and the *medians* of this affine triangle are the lines joining $\mathbf{a}$ to $\mathsf{midpoint}(\mathbf{b}, \mathbf{c})$, $\mathbf{b}$ to $\mathsf{midpoint}(\mathbf{a}, \mathbf{c})$, and $\mathbf{c}$ to $\mathsf{midpoint}(\mathbf{a}, \mathbf{b})$.

THEOREM II.25. *Let V, $\mathbb{F}$, $\mathbf{a},\ \mathbf{b},\ \mathbf{c}$ be as above. Then the medians of the affine triangle $\Delta\,\mathbf{abc}$ are concurrent (pass through a single point) if $1+1+1 \neq 0$ and parallel in pairs if $1+1+1 = 0$.*

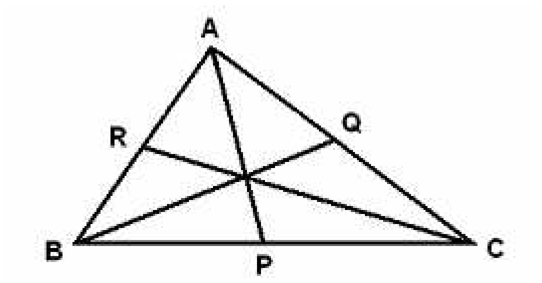


Figure II.4

Proof. *First case.* Suppose that $1+1+1 \neq 0$, and let $\frac{1}{3} = (1+1+1)^{-1}$. Assume that the point $\mathbf{x}$ lies on the line joining $\mathbf{a}$ to $\frac{1}{2}(\mathbf{b}+\mathbf{c})$ and also on the line joining $\mathbf{b}$ to $\frac{1}{2}(\mathbf{a}+\mathbf{c})$. Then there exist $s, t \in \mathbb{F}$ such that

$$s\mathbf{a} \ + \ (1-s)\tfrac{1}{2}(\mathbf{b}+\mathbf{c}) \ = \ \mathbf{x} \ = \ t\mathbf{b} \ + \ (1-t)\tfrac{1}{2}(\mathbf{a}+\mathbf{c}) \ .$$

Since $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are affinely independent in both expansions for $\mathbf{x}$ the coefficients of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ add up to 1, we may equate the barycentric coordinates in the two expansions for $\mathbf{x}$. In particular, this implies $s = \frac{1}{2}(1-t)$ and $t = \frac{1}{2}(1-s)$. If we solve these equations, we find that $s = t = \frac{1}{3}$, so that

$$\mathbf{x} \ = \ \tfrac{1}{3}(\mathbf{a} \ + \ \mathbf{b} \ + \ \mathbf{c}) \ .$$

A routine computation shows that this point does lie on both lines.

In a similar fashion one can show that the lines joining $\mathbf{a}$ to $\frac{1}{2}(\mathbf{b}+\mathbf{c})$ and $\mathbf{c}$ to $\frac{1}{2}(\mathbf{a}+\mathbf{b})$ also meet at the point $\frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c})$, and therefore we conclude that the latter point lies on all three medians.■

Second case. Suppose that $1+1+1 = 0$; in this case it follows that $\frac{1}{2} = -1$. The line joining $\mathbf{a}$ to $\frac{1}{2}(\mathbf{b}+\mathbf{c})$ is then given by $\mathbf{a}+W$, where W is spanned by

$$\tfrac{1}{2}(\mathbf{b}+\mathbf{c}) \ - \ \mathbf{a} \ = \ -(\mathbf{a}+\mathbf{b}+\mathbf{c}) \ .$$

Similarly computations show that the other two medians are given by $\mathbf{b}+W$ and $\mathbf{c}+W$. To complete the proof, we need to show that no two of these lines are equal.

However, if, say, we had $\mathbf{b}+W = \mathbf{c}+W$ then it would follow that $\mathbf{c}$, $\frac{1}{2}(\mathbf{a}+\mathbf{b})$, $\mathbf{b}$, and $\frac{1}{2}(\mathbf{a}+\mathbf{c})$ would all be collinear. Since the line joining the second and third of these points contains $\mathbf{a}$ by construction, it would then follow that $\mathbf{a} \in \mathbf{bc}$, contradicting our noncollinearity assumption. Thus $\mathbf{b}+W \neq \mathbf{c}+W$; similar considerations show that $\mathbf{a}+W \neq \mathbf{c}+W$ and $\mathbf{a}+W \neq \mathbf{b}+W$, and therefore the three medians are distinct (and by the preceding paragraph are each parallel to each other).■

HYPOTHESIS FOR THE REMAINDER OF THIS SECTION. *For the rest of this section, the vector space V is assumed to be* **two-dimensional.**

Definition. Let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d} \in V$ be four ordered points, no three of which are collinear. If $\mathbf{ab}||\mathbf{cd}$ and $\mathbf{ad}||\mathbf{bc}$, we shall call the union

$$\mathbf{ab} \ \cup \ \mathbf{bc} \ \cup \ \mathbf{cd} \ \cup \ \mathbf{da}$$

the *affine parallelogram* determined by $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$, and we shall write it $\square\mathbf{abcd}$. The *diagonals* of the parallelogram are the lines $\mathbf{ac}$ and $\mathbf{bd}$.

THEOREM II.26. *Suppose that* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d} \in V$ *as above are the vertices of an affine parallelogram. Then the diagonals* $\mathbf{ad}$ *and* $\mathbf{bc}$ *have a point in common and this point is equal to* $\mathsf{midpoint}(\mathbf{a}, \mathbf{d})$ *and* $\mathsf{midpoint}(\mathbf{b}, \mathbf{c})$.

If we take $\mathbb{F} = \mathbb{R}$ and $V = \mathbb{R}^2$, this result reduces to the classical geometric result that *the diagonals of a parallelogram bisect each other.*

Proof. The classical *parallelogram law* for vector addition states that

$$(\mathbf{c} - \mathbf{a}) \;=\; (\mathbf{d} - \mathbf{a}) \;+\; (\mathbf{b} - \mathbf{a})$$

(see the accompanying figure).

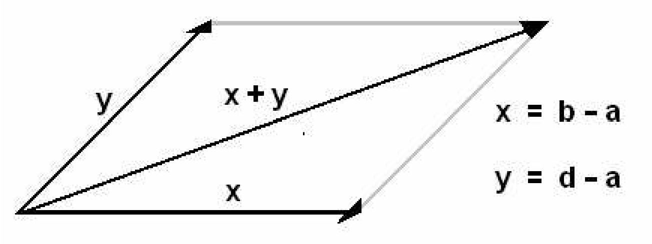


Figure II.5

In fact, it is trivial to verify that $\mathbf{d} + \mathbf{b} - \mathbf{a}$ lies on both the parallel to $\mathbf{ad}$ through $\mathbf{b}$ and the parallel to $\mathbf{ab}$ and $\mathbf{c}$, and hence this point must be $\mathbf{c}$. Therefore the common point of $\mathbf{ac}$ and $\mathbf{bd}$ satisfies the equations

$$t\mathbf{b} \;+\; (1-t)\mathbf{d} \;=\; s\mathbf{c} \;+\; (1-s)\mathbf{a} \;=\; s(\mathbf{d}+\mathbf{b}-\mathbf{a}) \;+\; (1-s)\mathbf{a} \;=\; s\mathbf{d} \;+\; s\mathbf{b} + (1-2s)\mathbf{a}\ .$$

Since $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{d}$ are affinely independent, if we equate barycentric coordinates we find that $s = t = \frac{1}{2}$. But this implies that the two lines meet at a point which is equal to both $\mathsf{midpoint}(\mathbf{a}.\mathbf{c})$ and $\mathsf{midpoint}(\mathbf{b}, \mathbf{d})$.■

We include the following two theorems because they help motivate the construction of projective space in Chapter III.

THEOREM II.27. *Let* $\mathbf{v}$, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ *be four points, no three of which are collinear (but coplanar). Let* $\mathbf{a}' \in \mathbf{va}$, $\mathbf{b}' \in \mathbf{vb}$, *and* $\mathbf{c}' \in \mathbf{vc}$ *be distinct from* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ *such that* $\mathbf{ab}||\mathbf{a}'\mathbf{b}'$ *and* $\mathbf{ac}||\mathbf{a}'\mathbf{c}'$. *Then* $\mathbf{bc}||\mathbf{b}'\mathbf{c}'$.

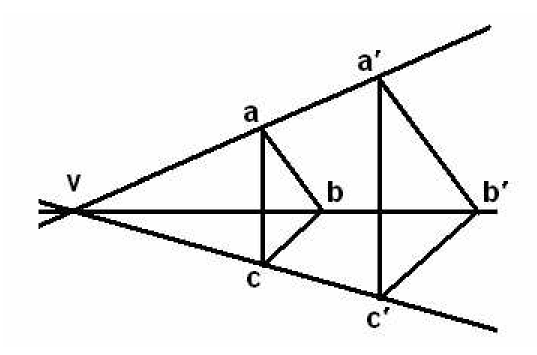


Figure II.6

Proof. Since $\mathbf{a}' \in \mathbf{va}$ and $\mathbf{b}' \in \mathbf{vb}$, we may write $\mathbf{a}' - \mathbf{v} + s(\mathbf{a} - \mathbf{v})$ and $\mathbf{b}' - \mathbf{v} + t(\mathbf{b} - \mathbf{v})$ for suitable scalars s and t. Since $\mathbf{ab}||\mathbf{a}'\mathbf{b}'$, it follows that $\mathbf{b}' - \mathbf{a}' = k(\mathbf{b} - \mathbf{a})$ for some scalar k. But

$$k(\mathbf{b}-\mathbf{v}) \; - \; k(\mathbf{a}-\mathbf{v}) \; = \; k(\mathbf{b}-\mathbf{a}) \; = \; \mathbf{b}'-\mathbf{a}' \; = \; (\mathbf{b}'-\mathbf{v}) \; - \; (\mathbf{a}'-\mathbf{v}) \; = \; t(\mathbf{b}-\mathbf{v}) \; - \; s(\mathbf{a}-\mathbf{v}) \; .$$

Since $\mathbf{b} - \mathbf{v}$ and $\mathbf{a} - \mathbf{v}$ are linearly independent, it follows that $s = t = k$. For similar reasons we also have $\mathbf{c}' - \mathbf{v} = t(\mathbf{c} - \mathbf{v})$.

To prove $\mathbf{bc}||\mathbf{bc}'$, it suffices to note that

$$\mathbf{c}' - \mathbf{b}' \; = \; (\mathbf{c}' - \mathbf{v}) \; - \; (\mathbf{b}' - \mathbf{v}) \; = \; t(\mathbf{c} - \mathbf{v}) \; - \; t(\mathbf{b} - \mathbf{v}) \; = \; t(\mathbf{c} - \mathbf{b})$$

by the preceding paragraph, and hence $\mathbf{bc}||\mathbf{b}'\mathbf{c}'$ follows.■

Here is a similar result with slightly different hypotheses:

Theorem II.28. *Let* $\mathbf{v}$, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ *be four points, no three of which are collinear (but coplanar). Let* $\mathbf{a}' \in \mathbf{va}$, $\mathbf{b}' \in \mathbf{vb}$, *and* $\mathbf{c}' \in \mathbf{vc}$ *be distinct from* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ *such that* $\mathbf{ab}||\mathbf{a}'\mathbf{b}'$ *but* $\mathbf{ac}$ *and* $\mathbf{a}'\mathbf{c}'$ *meet in some point* $\mathbf{x}$. *Then* $\mathbf{bc}$ *and* $\mathbf{b}'\mathbf{c}'$ *also meet in some point* $\mathbf{y}$ *and we have* $\mathbf{xy}||\mathbf{ab}$.

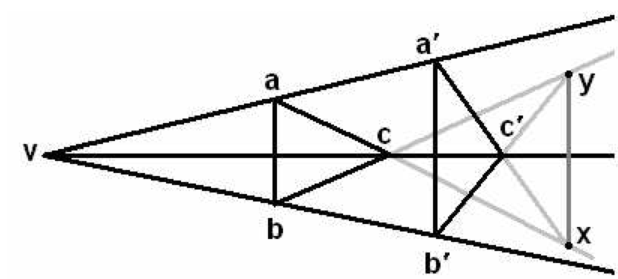


Figure II.7

Proof. As in Theorem 27 we may write $\mathbf{a}' - \mathbf{v}' = t(\mathbf{a} - \mathbf{v})$ and $\mathbf{b}' - \mathbf{v}' = t(\mathbf{b} - \mathbf{v})$; however, $\mathbf{c}' - \mathbf{v}' = s(\mathbf{c} - \mathbf{v})$ for some $s \neq t$ (otherwise $\mathbf{ac}||\mathbf{a}'\mathbf{c}'$). Expressions for the point $\mathbf{x}$ may be computed starting with the equation

$$\mathbf{x} \; = \; r\mathbf{a} \; + \; (1-r)\mathbf{c} \; = \; q\mathbf{a}' \; + \; (1-q)\mathbf{c}'$$

which immediately implies that

$$\mathbf{x}-\mathbf{v} \; = \; r(\mathbf{a}-\mathbf{v}) \; + \; (1-r)(\mathbf{c}-\mathbf{v}) \; = \; q(\mathbf{a}'-\mathbf{v}) \; + \; (1-q)(\mathbf{c}'-\mathbf{v}) \; = \; qt(\mathbf{a}-\mathbf{v}) \; + \; (1-q)s(\mathbf{c}-\mathbf{v}) \; .$$

Since $\mathbf{a} - \mathbf{v}$ and $\mathbf{c} - \mathbf{v}$ are linearly independent, we find that $r = qt$ and $1 - r = (1 - q)s$. These equations determine r completely as a function of s and t:

$$r(s,t) \; = \; \frac{t(1-s)}{s-t}$$

A similar calculation shows that any common point to $\mathbf{bc}$ and $\mathbf{b}'\mathbf{c}'$ has the form

$$r(s,t)\mathbf{b} \; + \; \big(1 - r(s,t)\big)\,\mathbf{c}$$

and a reversal of the previous argument shows that this point is common to $\mathbf{bc}$ and $\mathbf{b}'\mathbf{c}'$. Therefore

$$\mathbf{y} - \mathbf{x} \; = \; r(s,t)\,(\mathbf{b} - \mathbf{a})$$

which shows that $\mathbf{xy}||\mathbf{ab}$.■

There is a third result of a similar type; its proof is left as an exercise (we should note that all these results will be improved upon later).

THEOREM II.29. *In the notation of the theorems, assume that all three pairs of lines* $\{\,\mathbf{ab},\ \mathbf{a}'\mathbf{b}'\,\}$, $\{\,\mathbf{ac},\ \mathbf{a}'\mathbf{c}'\,\}$, *and* $\{\,\mathbf{bc},\ \mathbf{b}'\mathbf{c}'\,\}$ *all have points of intersection. Then the three intersection points are collinear.*

EXERCISES

Definition. Let $\mathbf{a},\ \mathbf{b},\ \mathbf{c},\ \mathbf{d} \in V$ be four ordered points, no three of which are collinear (with no parallelism assumptions). The union

$$\mathbf{ab} \ \cup\ \mathbf{bc}\ \cup\ \mathbf{cd}\ \cup\ \mathbf{da}$$

is called the *affine quadrilateral* determined by $= \mathbf{a},\ \mathbf{b},\ \mathbf{c}$ and $\mathbf{d}$, and as before we shall write it $\Box\mathbf{abcd}$. The *diagonals* of the quadrilateral are the lines **ac** and **bd**. The *sides* of the quadrilateral are four lines whose union forms the affine quadrilateral;

> *In the exercises below, assume that all vectors lie in the vector space* $\mathbb{F}^2$, *where* $\mathbb{F}$ *is a field in which* $1 + 1 \neq 0$.

1. Prove that an affine quadrilateral is a parallelogram if and only if its diagonals bisect each other (in the sense of Theorem 26).

2. Suppose we are given an affine parallelogram. Prove that a line joining the midpoints of a pair of parallel sides contains the intersection point of the diagonals.

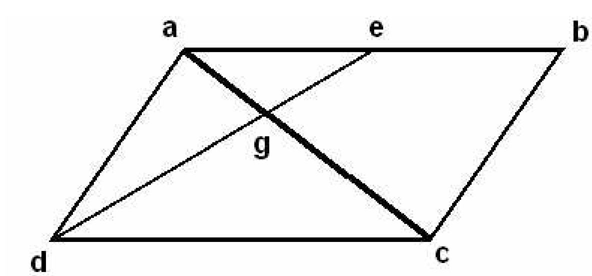


Figure II.8

3. In the figure above, assume we are given a parallelogram $\Box\mathbf{abcd}$ such that

$$\mathbf{e} \ = \ \mathsf{midpoint}(\mathbf{a},\mathbf{b})$$

and $1 + 1 + 1 \neq 0$ in $\mathbb{F}$. Prove that

$$\mathbf{g} \ = \ \tfrac{1}{3}\mathbf{c} \ + \ \tfrac{2}{3}\mathbf{a} \ = \ \tfrac{1}{3}\mathbf{c} \ + \ \tfrac{2}{3}\mathbf{e}\ .$$

Definition. An affine quadrilateral $\Box\mathbf{abcd}$ is said to be an *affine trapezoid* if either $\mathbf{ab}||\mathbf{cd}$ or $\mathbf{bc}||\mathbf{ad}$ but not both (and generally the points are labeled so that the first is true). The two parallel sides are called the *bases.*

4. Suppose we are given affine trapezoid $\square\mathbf{abcd}$ as above with bases $\mathbf{ab}$ and $\mathbf{cd}$. Prove that the line joining $\mathsf{midpoint}(\mathbf{a},\mathbf{d})$ and $\mathsf{midpoint}(\mathbf{b},\mathbf{c})$ is parallel to the bases.

5. In the same setting as in the previous exercise, prove that the line joining $\mathsf{midpoint}(\mathbf{a},\mathbf{c})$ and $\mathsf{midpoint}(\mathbf{b},\mathbf{d})$ is parallel to the bases.

6. In the same setting as in the previous exercises, prove that the line joining $\mathsf{midpoint}(\mathbf{a},\mathbf{d})$ and $\mathsf{midpoint}(\mathbf{b},\mathbf{c})$ is equal to the line joining $\mathsf{midpoint}(\mathbf{a},\mathbf{c})$ and $\mathsf{midpoint}(\mathbf{b},\mathbf{d})$.

7. Prove Theorem 29. [*Hint:* This can be done using the Theorem of Menelaus.]

In the next exercise, assume that all vectors lie in a vector space V over a field $\mathbb{F}$ in which $1+1=0$; the most basic example is the field $\mathbb{Z}_2$ which has exactly two elements.

8. Suppose that $1+1=0$ in $\mathbb{F}$, and aside from this we are in the setting of Theorem II.26: Specifically, let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}\in V$ be four noncollinear points such that the first three are noncollinear the four points are the vertices an affine parallelogram. Prove that in this case *the diagonal lines* $\mathbf{ad}$ *and* $\mathbf{bc}$ *are* **parallel**. Is the converse also true? Give reasons for your answer.

5. Generalized geometrical incidence

> Our geometry is an abstract geometry. The reasoning could be followed by a disembodied spirit who had no understanding of a physical point, just as a man blind from birth could understand the Electromagnetic Theory of Light. — **H. G. Forder (1889–1981)**

Although we have not yet defined a geometrical incidence space of arbitrary dimension, it is clear that the families of k-planes in $\mathbb{F}^n$ should define an n-dimensional incidence space structure on $\mathbb{R}^n$. Given this, it is not difficult to guess what the correct definitions should be.

Definition. A *geometrical incidence space* is a triple (S, Π, d) consisting of a set S, a family of subsets Π the *geometrical subspaces*), and a function d from Π to the positive integers (the *dimension*) satisfying the following conditions:

(G-1) : If $x_0 \; \cdots \; , x_n$ are distinct points of S such that no $P \in \Pi$ with $d(P) < n$ contains them all, then there is a unique $P \in \Pi$ such that $d(P) = n$ and $x_i \in P$ for all i.

Notation. We denote the geometrical subspace P in the preceding statement by $x_0 \; \cdots \; x_n$. The condition on the x_i is expressed in the statement that *the set* $\{ x_0 \; \cdots \; , x_n \}$ *is (geometrically) independent*.

(G-2) : If $P \in \Pi$ and $\{ x_0 \; \cdots \; , x_m \}$ is a set of geometrically independent points in P, then the geometrical subspace $x_0 \; \cdots \; x_m$ is contained in P.

(G-3) : If $P \in \Pi$, then P contains at least $d(P) + 1$ points.

If $P \in \Pi$ and $d(P) = k$, then we shall say that P is a k-plane; the set of all k-planes is denoted by Π_k. By convention, a 1-plane is often called a *line* and a 2-plane is often simply called a *plane*. Note that the defining conditions do not give any *a priori* information about whether or not there are any k-planes in S (however, if S contains at least two elements, one can prove that Π_1 must be nonempty).

For most of the examples that we shall consider, the whole space S is one of the geometrical subspaces. If this is the case and $\dim S = n$, then we shall say that the system (S, Π, d) is an *abstract n-dimensional geometrical incidence space*. When $n = 2$ we shall also say the system is an *abstract incidence plane*.

If we are given a geometrical incidence space and the explicit data Π and d are either clear from the context or are not really needed in a discussion, we shall often simply say that "S is a geometrical incidence space."

EXAMPLE 1. A three-dimensional incidence space (as defined above) is a geometrical incidence space if we take $\Pi_1 = \mathcal{L}$, $\Pi_2 = \mathcal{P}$, and $\Pi_3 = \{S\}$.

EXAMPLE 2. The **minimal** geometrical indicidence structures: Let Π be the family of all finite subsets of a set S with at least two elements, and for each such subset Q let $d(Q) = \#(Q) - 1$, where $\#(Q)$ is the number of elements in Q. The figure below illustrates the special case in which S has three elements $\{\mathbf{a},\ \mathbf{b},\ \mathbf{c}\}$ and $d(S) = 2$.

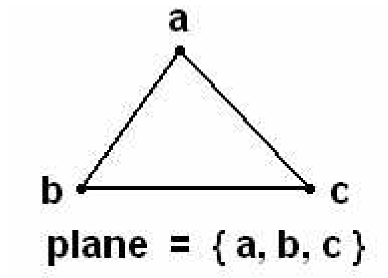


Figure II.9

EXAMPLE 3. An important class of examples mentioned at the beginning of this section: Let V be an n-dimensional vector space over a field $\mathbb{F}$ (where $n > 0$), and define the *affine incidence structure on* V such that for each positive integer k the set Π_k is the set of k-planes considered in earlier sections of this chapter. The properties **(G-1)**–**(G-3)** may be verified as follows:

Proof of **(G-1)**. If a set of vectors $\{ \mathbf{v}_0, \cdots, \mathbf{v}_k \}$ is not geometrically independent, then the set is also affinely dependent, for if $\{ \mathbf{v}_0, \cdots, \mathbf{v}_k \}$ is contained in a q-plane $\mathbf{y} + W$ for some $q < k$, then we have $\mathbf{y} + W = \mathbf{x}_0 + W$ and the k vectors $\mathbf{v}_1 - \mathbf{v}_0, \cdots, \mathbf{v}_k - \mathbf{v}_0 \in W$ must be linearly dependent because $\dim W < k$. Hence the original vectors are affinely dependent as claimed. Taking negations, we see that if $\{ \mathbf{v}_0, \cdots, \mathbf{v}_k \}$ is geometrically independent, then the set is also affinely independent.

Let Q be the affine span of $\mathbf{v}_0, \cdots, \mathbf{v}_k$; then $Q = bfx_0 + W$ where W is the linear span of $\mathbf{v}_1 - \mathbf{v}_0, \cdots, \mathbf{v}_k - \mathbf{v}_0$, and W is k-dimensional because these vectors are linearly independent. Therefore Q is a k-plane containing the vectors $\mathbf{v}_i$. Conversely, if Q' is an arbitrary k-plane containing the vectors $\mathbf{v}_i$, then we may write $Q' = \mathbf{v}_0 + U$ where U is a vector subspace of dimension k which contains the difference vectors $\mathbf{v}_1 - \mathbf{v}_0, \cdots, \mathbf{v}_k - \mathbf{v}_0$; it follows that U contains the k-dimensional vector subspace W described above, and since $\dim U = \dim W$ it follows that $W = U$, so that $= Q = Q'$.■

Proof of **(G-2)**. Since a k-plane is closed under forming affine combinations, if $\mathbf{v}_0, \cdots, \mathbf{v}_m$ is contained in P then the affine span of $\mathbf{v}_0, \cdots, \mathbf{v}_m$ is also contained in P.■

Proof of **(G-3)**. Given a k-plane P, express it as $\mathbf{v}_0 + W$, where W is a k-dimensional vector subspace, and let $\mathbf{v}_0, \cdots, \mathbf{v}_k$ be an affine basis for W. Then the set

$$\{ \mathbf{v}_0, \ \mathbf{v}_0 + \mathbf{v}_1, \ \cdots, \mathbf{v}_k + \mathbf{v}_1 \}$$

forms an affine basis for P by Theorem 22. Hence P contains at least $(k+1)$ points.■

We have not yet described an analog of the axiom implying that two planes in 3-space intersect in a line if they are not disjoint. Before formulating an appropriate generalization of this,[6] we derive some consequences of **(G-1)**–**(G-3)**:

THEOREM II.30. *Let $x_0, \cdots, x_m$ be geometrically independent points in a geometrical incidence space S, and suppose $y \not\in \{x_0, \cdots, x_m\}$. Then the set $\{x_0, \cdots, x_m, y\}$ is geometrically independent.*

[6]The definition follows the proof of Theorem 35 below.

Proof. If the points are not geometrically independent then for some $k \leq m$ there is a k-plane P which contains all of them. Since $\{x_0, \cdots, x_m\}$ is geometrically independent, it follows that $d(P) \geq m$, so that $d(P) = m$ and $P = x_0 \ \cdots \ x_m$. But this contradicts the hypothesis.■

THEOREM II.31. *Let P be a k-plane in the geometrical incidence space S. Then there is a set of independent points $\{x_0, \cdots, x_m\}$ such that $P = x_0 \ \cdots \ x_m$.*

Proof. Let $\mathcal{F}$ be the family of finite independent subsets of P. No element in this family contains more than $m+1$ elements, for every subset with more points will be geometrically dependent. Let k be the largest integer for which some element of $\mathcal{F}$ has $k+1$ elements; by the preceding sentence, we must have $k \leq m$. By **(G-1)** we have $k = m$.

Assume the contrary, so that $k < m$. Let $\{x_0, \cdots, x_k\} \in \mathcal{F}$. Then the k-plane $Q = x_0 \ \cdots \ x_k$ is contained in $\mathcal{F}$, and $k < m$ implies that Q is a proper subset of P. Let x_{k+1} be a point which is in P but not in Q. Then Theorem 30 implies that $\{x_0, \cdots, x_k, y\}$ is geometrically independent. But this contradicts the defining condition for k, which is that there are no geometrically independent subsets of P with $k+2$ elements. Therefore we must have $k = m$.■

STANDARD CONVENTIONS. Given a geometrical incidence space (S, Π, d), we shall often say that the empty set is a geometrical subspace whose dimension is -1, each one point subset of S is a geometrical subspace whose dimension is zero, and subsets of both these types are geometrically independent. — Properties **G-1)** – **(G-3)** of a geometrical incidence space remain valid if we extend the definition of geometrical subspace using these conventions, and the latter allow us to state the next result more simply.

THEOREM II.32. *Every finite subset X of the geometrical incidence space S contains a maximal independent subset $Y = \{y_0, \cdots, y_k\}$. Furthermore, X is contained in $y_0 \ \cdots \ y_k$, and the latter is the unique minimal geometrical subspace containing X.*

Proof. Let $Y \subset X$ be an independent subset with a maximum number of elements, let Q be the k-plane determined by Y, and let $w \in X$ be an arbitrary element not in Y. Since $y \notin Q$ would imply that $Y \cup \{y\}$ would be independent, it follows that $y \in Q$. Thus $X \subset Q$ as claimed. Suppose that Q' is another geometrical subspace containing X; then $Q \subset Q'$ by **(G-2)**, and hence every geometrical subspace that contains X must also contain Q.■

When we work with vector subspaces of a vector space, it is often useful to deal with their intersections and sums. The next two results show that similar constructions hold for geometrical subspaces:

THEOREM II.33. *The intersection of a nonempty family of geometrical subspaces of S is a geometrical subspace (with the conventions for 0- and (-1)-dimensional subspaces preceding Theorem 32).*

Proof. Clearly it suffices to consider the case where the intersection contains at least two points. Let $\{P_\alpha \mid \alpha \in A\}$ be the family of subspaces, and let $\mathcal{F}$ be the set of all finite independent subsets of $\cap_\alpha P_\alpha$. As before, the number of elements in a member of $\mathcal{F}$ is at most $d(P_\alpha) + 1$ for all α. Thus there is a member of $\mathcal{F}$ with a maximal number of elements; call this subset $\{x_0, \cdots, x_k\}$. By **(G-2)** we know that $x_0 \ \cdots \ x_k$ is contained in each P_α and hence in $\cap_\alpha P_\alpha$. If it were a proper subset and y is a point in the intersection which does not belong to $x_0 \ \cdots \ x_k$, then $\{x_0 \ \cdots \ x_k, y\}$ would be an independent subset of the intersection with more elements than $\{x_0, \ \cdots \ , x_k\}$. This contradiction means that $x_0 \ \cdots \ x_k = \cap_\alpha \ P_\alpha$.■

Although the union of two geometrical subspaces P and Q is usually not a geometrical subspace, the next result shows that there is always a minimal geometrical subspace containing both P and Q; this is analogous to the concept of sum for vector subspaces of a vector space.

THEOREM II.34. *If P and Q are geometrical subspaces of the geometrical incidence space S, then there is a unique minimal geometrical subspace containing them both.*

Proof. By Theorem 31 we may write $P = x_0 \ \cdots \ x_m$ and $Q = y_0 \ \cdots \ y_n$. Let $A = \{x_0, \ \cdots \ , x_m, y_0, \ \cdots \ , y_n\}$, and let T be the smallest subspace containing A (which exists by Theorem 32). Then $P, Q \subset T$, and if T' is an arbitrary geometrical subspace containing P and Q then it also contains A, so that T' must contain T as well.■

The subspace given in the preceding result is called the **join** of P and Q, and it is denoted by $P \star Q$. Motivation for this definition is given by Exercise III.4.17 and Appendix B.

THEOREM II.35. *If P and Q are geometrical subspaces of S, then $d(P \star Q) \leq d(P) + d(Q) - d(P \cap Q)$.*

It is easy to find examples of geometrical subspaces in $\mathbb{R}^n$ (even for $n = 2$ or 3) in which one has **strict** inequality. For example, suppose that L and M are parallel lines in $\mathbb{R}^3$; then the left hand side is equal to 2 but the right hand side is equal to 3 (recall that the empty set is (-1)-dimensional). Similarly, one has strict inequality in $\mathbb{R}^3$ if P is a plane and Q is a line or plane which is parallel to P.

Proof. Let $P \cap Q = x_0 \ \cdots \ x_m$. By a generalization of the argument proving Theorem 31 (see Exercise 1 below), there exist independent points $y_0, \ \cdots \ , y_p \in P$ and $z_0, \ \cdots \ , z_q \in Q$ such that

$$P \ = \ x_0 \ \cdots \ x_m y_0 \ \cdots \ y_p \ , \qquad Q \ = \ x_0 \ \cdots \ x_m z_0 \ \cdots \ z_q \ .$$

Let $X = \{x_0, \ \cdots \ , \ x_m, y_1, \ \cdots \ , y_p, z_1, \ \cdots \ , z_q$, and let T be the unique smallest geometrical subspace containing X. It is immediate that $P \subset T$ and $Q \subset T$, so that $P \star Q \subset T$. On the other hand, if a geometric subspace B contains $P \star Q$, then it automatically contains X and hence automatically contains T. Therefore we have $T = P \star Q$.

It follows from Theorem 32 that $d(P \star Q) = \dim S \leq \#(X) + 1$, and therefore we have

$$\dim(P \star Q) \ \leq \ m \ + \ 1 \ + \ p \ + \ q \ = \ (m+p+1) \ + \ (m+q+1) \ - \ (m+1) \ = \\ d(P) \ + \ d(Q) \ - \ d(P \cap Q)$$

which is what we wanted to prove.■

The following definition contains an abstract version of the 3-dimensional incidence assumption about two planes intersecting in a line or not at all.

Definition. A geometrical incidence space is *regular* if the following holds:

(G-4) : If P and Q are geometrical subspaces such that $P \cap Q \neq \varnothing$, then

$$d(P \star Q) \ = \ d(P) \ + \ d(Q) \ - \ d(P \cap Q) \ .$$

EXAMPLE 1. Ordinary 3-dimensional incidences as defined in Section II.1 are regular. The only nontrivial case of the formula arises when P and Q are distinct planes, so that $P \star Q = S$.

EXAMPLE 2. The minimal examples at the beginning of this section are regular. The formula in this case follows immediately from the standard *inclusion-exclusion identity* for counting the elements in finite sets:

$$\#(P \cup Q) \ = \ \#(P) \ + \ \#(PQ) \ - \ \#(P \cap Q) \ .$$

EXAMPLE 3. Here is an example which is **not** regular. Let $\mathbb{F}$ be a field, take the standard notions of lines and planes for $\mathbb{F}^4$, and let $d(\mathbb{F}^4) = 3$. Then the planes V and W spanned by $\{\mathbf{e}_1, \mathbf{e}_2\}$ and $\{\mathbf{e}_3, \mathbf{e}_4\}$ have exactly one point in common.

Logical indpendence of the regularity condition. The preceding example shows that it is **not** possible to prove the assumption in the definition of regularity from the defining assumptions for a geometrical incidence space. — For if it were possible to prove the regularity condition from the definition, then it would NOT be possible to construct an example of a geometrical incidence space that did not satisfy the regularity condition.

The preceding remark illustrates the mathematical approach to concluding that one statement (say $\mathbf{Q}$) cannot be derived as a logical consequence of other statements (say $\mathbf{P}_1, \cdots, \mathbf{P}_n$): It is only necessary to produce and example of an object which satisfies $\mathbf{P}_1, \cdots, \mathbf{P}_n$ but does not satisfy $\mathbf{Q}$.

EXAMPLE 4. The incidence structure associated to a finite-dimensional vector space V over a field $\mathbb{F}$ is regular. To prove this, we first establish the following.

THEOREM II.36. *Let P and Q be affine subspaces of the incidence space structure associated to a vector space V, and assume $P \cap Q \neq \varnothing$. Write $P = \mathbf{x} + W$ and $Q = \mathbf{x} + U$ where $\mathbf{x} \in P \cap Q$. Then $P \star Q = \mathbf{x} + (W_1 + W_2)$.*

REMARK. Since $P \cap Q = \mathbf{x} + (W_1 \cap W_2)$ is readily established (see Exercise II.2.4), Theorem 36 and the dimension formula for vector subspaces (Theorem A.9) imply the regularity of V.■

Proof. The inclusion $P \star Q \subset \mathbf{x} + (W_1 + W_2)$ is clear since the right hand side is a geometrical subspace containing P and Q. To see the reverse inclusion, first observe that $P \star Q = \mathbf{x} + U$ for some vector subspace U; since $P,\ Q \subset \mathbf{x} + U$ it follows that $W_1,\ W_2 \subset U$ and hence $W_1 + W_2 \subset U$. The latter yields $\mathbf{x} + (W_1 + W_2) \subset P \star Q$, and therefore the two subsets are equal. ■

Finally, we introduce an assumption reflecting the Euclidean Parallel Postulate.

Definition. Two lines in a geometrical incidence space are *parallel* if they are coplanar but disjoint. A regular incidence space is said to be affine if given a line L and a point $\mathbf{x} \notin L$, then there is a unique line M such that $\mathbf{x} \in M$ and M is parallel to L. — If S is an affine incidence space and $d(S) = n$, then we say that it is an *affine n-space*. If $n = 2$, it is also called an *affine plane*.

One can use the argument proving Theorem 14 to verify that the affine incidence space associated to a vector space is affine in the sense defined above.

Equivalent mathematical structures

Our discussion of geometrical incidence fits closely with the main themes of these notes, but the formulation is definitely nonstandard. Normally such incidence structures are viewed in equivalent but more abstract terms. One approach involves specifying a sequence of *dependence relations* on finite subsets of a given set S; in this formulation, there is a class of such subsets that are called *independent* and satisfy a few suitable properties. As one might expect, a finite subset of $k+1$ points is said to be independent if and only if there is no k-dimensional subspace containing them; the resulting structure is called a **matroid**. Details of abstract dependence theory and matroids are described in the paper by H. Whitney listed in the bibliography, and a more recent account of matroid theory is given in the following online reference:

`http://home.gwu.edu/~jbonin/survey.ps`

The matroid approach to independence leads naturally to another interpretation in terms of partially ordering relations on sets. Every matroid has a family of subsets which satisfy the conditions for a geometrical incidence space, and the associated ordered family of subsets satisfies the conditions for a partially ordered set to be a **geometric lattice**. A classic reference for the theory of lattices is the book by G. Birkhoff (*Lattice Theory*) cited in the bibliography.

EXERCISES

1. Let P be a geometrical subspace of S, and suppose that $\{x_0, \cdots, x_k\}$ is an independent subset of P. Prove that there is a (possibly empty) set of points $x_{k+1}, \cdots, x_m \in P$ such that $\{x_0, \cdots, x_m\}$ is independent and $P = x_0 \cdots x_m$. [*Hint:* Imitate the proof of Theorem 31 using the family $\mathcal{G} \subset \mathcal{F}$ of all subsets containing $\{x_0, \cdots, x_k\}$.]

2. Prove that a subset of an independent set of points in a geometrical incidence space is independent.

3. (i) Let (S, Π, d) be a geometrical incidence space, let $T \subset S$ be a geometrical subspace, and let Π_T be the set of all geometrical subspaces in Π which are contained in T. Prove that (S, Π_T, d_T) is a geometrical incidence space, where d_T is the restriction or T.

(ii) Prove that a geometrical subspace of a regular geometrical incidence space is again regular. Is the analog true for affine spaces? Give reasons for your answer.

4. Let S be a geometrical incidence space with $d(S) = n$. A *hyperplane* in S is an $(n-1)$-plane. Prove that for every $k < n$, every k-plane in S is an intersection of $(n-k)$ distinct hyperplanes. [*Hint:* If T is a k-plane and $T = x_0 \cdots x_k$, choose $y_1, \cdots, y_{n-k}$ such that $\{x_0 \cdots x_k, y_1, \cdots, y_{n-k}\}$ is independent. Let P_i be the hyperplane determined by all these points except y_j.]

5. Prove that the join construction on geometrical subspaces of a geometrical incidence space satisfies the associativity condition $(P \star Q) \star R = P \star (Q \star R)$ for geometrical subspaces $P.\ Q,\ R \subset S$. [*Hint:* Show that both of the subspaces in the equation are equal to the smallest geometrical subspace containing $P \cup Q \cup R$.]

6. Show that the Euclidean parallelism property is not a logical consequence of the defining conditions for a regular n-dimensional geometrical incidence space. [*Hint:* Look at the so-called trivial examples from the previous section in which Π is the collection of all finite subsets and if P is finite then $d(P) = \#(P) - 1$. Why do standard counting formulas imply that every pair of coplanar lines in this system will have a common point?]

Remark. For affine planes and the examples considered in the previous section, the following modification of the Euclidean Parallelism Property is valid: *Given a line L and a point $\mathbf{x} \notin L$, then there is* **at most one** *line M such that $\mathbf{x} \in M$ and M is parallel to L.* It is also not possible to prove this from the conditions defining a regular incidence space, and the following *Beltrami-Klein incidence plane* is an example: Let P be the set of all points in $\mathbb{R}^2$ which are interior to the unit circle; this set is defined by the inequality $x^2 + y^2 < 1$ (see the figure below). The lines in P are taken to be the open chords of the unit circle with equation $x^2 + y^2 = 1$, or equivalently all **nonempty** intersections of ordinary lines in $\mathbb{R}^2$ with P. As the figure suggests, this example has the following property: *Given a line L and a point $\mathbf{x} \notin L$, then there are* **at least two** *lines M such that $\mathbf{x} \in M$ and M is parallel to L.* — In fact, there are *infinitely many* such lines.

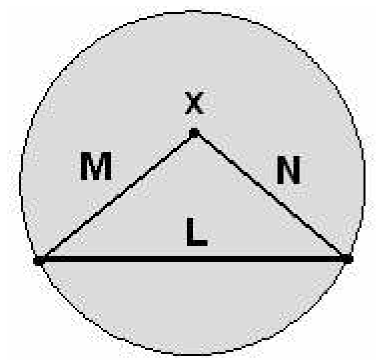


Figure II.10

The plane is the shaded region inside the circle, with the boundary excluded. The open chord L has no points in common with the open chords determined by M and N.

7. (*Further examples as above*) Define a *finitely supportable convex body* in $\mathbb{R}^n$ to be a nonempty set D defined by finitely many strict inequalities $f_i(\mathbf{x}) > 0$, where each f_i is a linear polynomial in the coordinates of $\mathbf{x}$. Let Π_D denote the set of all nonempty intersections $P \cap D$, where P is a geometrical subspace of $\mathbb{R}^n$, and set the dimension $d_D(P \cap D$ equal to $d(P)$. Prove that this structure makes D into a regular geometrical incidence n-space. [*Hint:* Try to do this first for $n = 2$, then for $n = 3$.]

8. Assume we are in the setting of the preceding exercise.

(*i*) Let D be the open square in $\mathbb{R}^2$ defined by $x > 0$, $y > 0$, $-x > -1$ and $-y > -1$ (hence the coordinates lie between 0 and 1). Prove that D has the parallelism properties that were stated but not proved for the Beltrami-Klein plane described above.

(*ii*) Let D be the upper half plane in $\mathbb{R}^2$ defined by $y > 0$. Prove that for some lines L and external points $\mathbf{x}$ there is a unique parallel to L through $\mathbf{x}$, but for others there are infinitely many such parallels.

6. Isomorphisms and automorphisms

Although the mathematical concepts of *isomorphism* and *automorphism* are stated abstractly, they reflect basic concepts that arise very quickly in elementary geometry. Congruent triangles are fundamental examples of isomorphic objects: The statement $\Delta\,ABC \cong \Delta\,DEF$ means that the obvious 1–1 correspondence of vertices from $\{A, B, C\}$ to $\{D, E, F\}$ preserves the basic structural measurements, so that the distances between the vertices satisfy

$$d(A,B) \;=\; d(D,E), \qquad d(B,C) \;=\; d(E,F), \qquad d(A,C) \;=\; d(D,F)$$

and the (degree or radian) measurements of the vertex angles satisfy similar equations:

$$\begin{aligned} \mathsf{measure}(\angle\,ABC) &= \mathsf{measure}(\angle\,DEF) \\ \mathsf{measure}(\angle\,ACB) &= \mathsf{measure}(\angle\,DFE) \\ \mathsf{measure}(\angle\,CAB) &= \mathsf{measure}(\angle\,FDE) \end{aligned}$$

Suppose now that we have an isosceles triangle $\Delta\,ABC$ in which $d(A,B) = d(A,C)$. In such cases one has a natural symmetry of the triangle with respect to the line joining A to the midpoint of B and C, and one aspect of this symmetry is a nontrivial congruence of the isosceles triangle with itself; specifically, $\Delta\,ABC \cong \Delta\,ACB$. In mathematics, an isomorphism from an object to itself is called an automorphism. The identity map from the object to itself is a trivial example of an automorphism, and the isosceles triangle example indicates that some structures may have other automorphisms.

The word "isomorphic" means "identical shape" or "same form," and in mathematics it means that one has a rule for passing between two objects that preserves all the mathematical structure that is essential in a given context. In particular, *if two objects are isomorphic, a statement about the structure of the first object is true if and only if the corresponding statement about the second object is true*. Such principles can be extremely useful, especially when one of the objects is relatively easy to work with and the other is less so.

The precise mathematical definitions of isomorphisms and automorphisms vary, and the details depend upon the sort of objects being considered. Also, there may be several different notions of isomorphism or automorphism, depending upon the amount of structure that remains unchanged. For example, if we are working with triangles, it is often useful to consider triangles that might not be congruent but are *similar*; in other words, one still has a 1–1 correspondence correspondence of vertices from $\{A, B, C\}$ to $\{D, E, F\}$ as before such that the angle measure equations are satisfied, but now we only know that there is a positive number r (the *ratio of similitude*) such that

$$d(A,B) \;=\; r \cdot d(D,E), \quad d(B,C) \;=\; r \cdot d(E,F), \quad d(A,C) \;=\; r \cdot d(D,F)\ .$$

Of course, congruence is the special case of similarity for which $r = 1$.

Here is the appropriate definition for geometrical incidence spaces:

Definition. Let (S, Π, d) and (S', Π', d') be geometrical incidence spaces. An *isomorphism* of geometrical incidence spaces from (S, Π, d) to (S', Π', d') is a 1–1 correspondence $f : S \to S'$ such that if $P \subset S$, then $P \in \Pi$ if and only if its image $f[P]$ belongs to Π', and in this case $d(P) = d'(\,f[P]\,)$. In other words, P is a k-plane in S if and only if $f[P]$ is a k-plane in S'.[7]

[7] As noted in the Prerequisites, we use $f[P]$ to denote the image of P under the mapping f.

It is standard to write $A \cong B$ when two mathematical systems are isomorphic, and we shall do so throughout these notes.

The first theorem of this section implies that isomorphism of geometrical incidence spaces is an equivalence relation.

THEOREM II.37. *(i) If $f : S \to S'$ is an isomorphism of geometrical incidence spaces, then so is its inverse $f^{-1} : S' \to S$.*

(ii) If $f : S \to S'$ and $g : S' \to S''$ are isomorphisms of geometrical incidence spaces, then so is their composite $g \circ f : S \to S''$.

Proof. (*i*) If f is 1–1 and onto, then it has an inverse map f^{-1} which is also 1–1 and onto. If $Q \subset S$ is a k-plane, then the identity $Q = f\left[\, f^{-1}[Q]\,\right]$ implies that $f^{-1}[Q]$ is a k-plane in S. Similarly, if $f^{-1}[Q]$ is a k-plane in S then so is $Q = f\left[\, f^{-1}[Q]\,\right]$.

(*ii*) If f and g are 1–1 onto, then so is $g \circ f$. If $P \subset S$ is a k-plane, then so is $f[P]$ because f is an isomorphism of geometrical incidence spaces, and since g is also an isomorphism of geometrical incidence spaces then

$$g \circ f[P] \;=\; g\big[\, f[P]\,\big]$$

is also a k-plane. Conversely, if the latter is a k-plane, then so is $f[P]$ since g is an isomorphism of geometrical incidence spaces, and therefore P is too because f is an isomorphism of geometrical incidence spaces.■

The next result illustrates earlier assertions that isomorphisms preserve the basic properties of mathematical structures like abstract geometrical incidence spaces.

THEOREM II.38. *Let $f : S \to S'$ is an isomorphism of geometrical incidence spaces, and let $X = \{\mathbf{x}_0, \;\cdots\; \mathbf{x}_m\}$ be a finite subset of S. Then X is a geometrically independent subset of S if and only if $f[X]$ is a geometrically independent subset of X'.*

Proof. Suppose that X is independent, and assume that $f[X]$ is not. Then there is some k-plane $Q \subset S'$ such that $d(P) < q$ and $f[X] \subset Q$. Let $P = f^{-1}[Q]$. By the definition of isomorphism, it follows that P is also a k-plane, where $k < q$, and $X = f^{-1}\big[\, f[X]\,\big]$ is contained in P. Therefore X is not independent, which contradicts our choice of X. Thus it follows that $f[X]$ is geometrically independent if X is. The converse statement follows by applying the preceding argument to the inverse isomorphism f^{-1}.■

If $\mathcal{G}$ is a family of geometrical incidence spaces, then a *classifying family* for $\mathcal{G}$ is a subfamily $\mathcal{C}$ such that each object in $\mathcal{G}$ is isomorphic to a unique space in $\mathcal{C}$. The standard coordinate affine spaces $\mathbb{F}^n$ (where $\mathbb{F}$ is some field) comes relatively closed to being a classifying family for n-dimensional affine incidence spaces for $n \geq 3$. It is only necessary to add standard coordinate affine n-spaces over *skew-fields* (see the second paragraph of Appendix A). The only difference between fields and skew-fields is that multiplication is not necessarily commutative in the latter; the standard nontrivial example is given by the **quaternions**, which are discussed in Appendix A. If $\mathbb{F}$ is a skew-field, let $\mathbb{F}^n$ be the **right** vector space of ordered n-tuples as defined in Appendix A. As indicated there, all of the linear algebra used to prove that $\mathbb{F}^n$ is an affine n-space goes through if $\mathbb{F}$ is a skew-field. In particular, the right vector space $\mathbb{F}^n$ is an affine n-space. The classification of n-dimensional affine spaces (where $n \geq 3$) is then expressible as follows:

THEOREM II.39. *Let $(S, etc.)$ be an affine n-space, where $n \geq 3$. Then there is a skew-field $\mathbb{F}$ such that S is isomorphic to $\mathbb{F}^n$ as a geometrical incidence space. Furthermore, if $\mathbb{E}$ ind $\mathbb{F}$ are skew-fields such that S is isomorphic to $\mathbb{E}^n$ and $\mathbb{F}^n$, then $\mathbb{E}$ and $\mathbb{F}$ are algebraically isomorphic.*

The proof of this result involves several concepts we have not yet introduced and is postponed to Remark 3 on following the proof of Theorem IV.19. Specifically, it will be a fairly simple application of the projective coordinatization theorem (Theorem IV.18) to the synthetic projective extension of S (which is defined in the Addendum to Chapter 3).

Automorphisms and symmetry

A standard dictionary definition of the word *symmetry* is "regularity in form and arrangement," and the significance of automorphisms in mathematics is that they often yield important patterns of regularity in a mathematical system. This is apparent in our previous example involving isosceles triangles. Of course, the amount of regularity can vary, and an equilateral triangle has a more regular structure than, say, an isosceles right triangle; generally speaking, the more regular an object is, the more automorphisms it has.

Definition. An *automorphism* of a geometrical incidence space (S, Π, d) and (S', Π', d') is a 1–1 onto map from S to itself which is an isomorphism of geometrical incidence spaces. — The identity map of a mathematical object is normally an example of an automorphism, and it is a routine exercise to check that for each geometrical incidence space (S, Π, d), the identity map of S defines an automorphism of (S, Π, d).

The following result is an immediate consequence of Theorem 38 and the preceding sentence.

THEOREM II.40. *The set of all automorphisms of a geometrical incidence space (S, Π, d) forms a group with respect to composition of mappings.*■

Notation. This group is called the *geometric symmetry group* of (S, Π, d) and is denoted by $\mathbf{\Sigma}(S, \Pi, d)$.

EXAMPLE 1. Let (S, Π, d) be the affine incidence space associated to a vector space V. If T is an invertible linear self-map of V, then T also defines an incidence space automorphism. This is true because the condition $\mathrm{Kernel}(T) = \{\mathbf{0}\}$ implies that T maps each k-dimensional vector subspace W to another subspace $T[W]$ of the same dimension (see Theorem A.14.(iv)). Furthermore, if S is an arbitrary subset and $\mathbf{x} \in V$, then it is easy to check that $T[\mathbf{x} + S] = T(\mathbf{x}) + T[S]$ (this is left to the reader as an exercise). .

EXAMPLE 2. If V is as above and $\mathbf{x} \in V$, define the mapping $T_{\mathbf{x}} : V \to V$ (*translation by* $\mathbf{x}$) to be the mapping $T_{\mathbf{x}}(\mathbf{v}) = \mathbf{x} + \mathbf{v}$. Then $T_{\mathbf{x}}$ is clearly 1–1 and onto. Furthermore, it defines a geometrical space automorphism because $T_{\mathbf{x}}[\mathbf{y} + W] = (\mathbf{x} + \mathbf{y}) + W$ shows that P is a k-plane in V if and only if $T_{\mathbf{x}}[P]$ is (again, filling in the details is left to the reader).■

The preceding observations show that affine incidence spaces associated to vector spaces generally have many geometric symmetries. Define the *affine group* of V — denoted by $\mathrm{Aff}(V)$ — to be all symmetries expressible as a composite $T \circ S$, where T is a translation and S is an invertible linear transformation. We shall say that the elements of $\mathrm{Aff}(V)$ are *affine transformations*. The terminology suggests that the affine transformations form a subgroup, and we shall verify this in the next result.

THEOREM II.41. *The set* $\mathrm{Aff}(V)$ *is a subgroup of the group of all geometric symmetries of* V. *It contains the groups of linear automorphisms and translations as subgroups.*

Proof. First of all, the identity map is a linear transformation and it is also translation by $\mathbf{0}$, so clearly the identity belongs to $\mathrm{Aff}(V)$. If A is an arbitrary affine transformation, then for each $\mathbf{x} \in V$ we have $A(\mathbf{x}) = S(\mathbf{x}) + \mathbf{y}$, where $\mathbf{y} \in V$ and S is an invertible transformation. If A' is another such transformation, write $A'(\mathbf{x}) = S'(\mathbf{x}) + \mathbf{y}'$ similarly. Then we have

$$A' \circ A(\mathbf{x}) \;=\; A'\big(\, S(\mathbf{x}) + \mathbf{y}\,\big) \;=$$
$$S'\big(\,(S(\mathbf{x}) + \mathbf{y})\,\big) + \mathbf{y}' \;=\; S'S(\mathbf{x}) + \big(\, S'(\mathbf{y}) + \mathbf{y}'\,\big)$$

showing that $A' \circ A$ is also affine; hence $\mathrm{Aff}(V)$ is closed under multiplication. To see that this family is closed under taking inverses, let A be as above, and consider the transformation

$$B(\mathbf{x}) \;=\; S^{-1}(\mathbf{x}) \;-\; S^{-1}(\mathbf{y}) \;.$$

By the above formula, $B \circ A$ and $A \circ B$ are both equal to the identity, so that A^{-1} is the affine transformation B. The second statement in the theorem follows from Examples 1 and 2 above.■

In general, $\mathrm{Aff}(V)$ is **not** the full group of geometric symmetries. For example, if $\mathbb{F} = \mathbb{C}$ (the complex numbers), the cooreinatewise conjugation map

$$\chi : \mathbb{C}^n \;\longrightarrow\; \mathbb{C}^n$$

taking $(z_1, \;\cdots\;, z_n)$ to $(\overline{z_1}, \;\cdots\;, \overline{z_n})$ is a geometrical symmetry of $\mathbb{C}^n$ which does not lie in $\mathrm{Aff}\,(\mathbb{C}^n)$.[8] However, if $\mathbb{F}$ is the integers mod p (where p is prime, the rational numbers, or the real numbers, then $\mathrm{Aff}(V)$ does turn out to be the entire group of geometric symmetries of $\mathbb{F}^n$.

Historical note. Although the concept of geometric symmetry was not formulated explicitly until the early 19th century, special symmetries of $\mathbb{R}^n$ (for at least $n = 2, 3$) known as *rigid motions* or *isometries* are implicit in classical Greek geometry, particularly in attempts at "proof by superposition" (a rigid motion is defined to be a 1–1 correspondence T from $\mathbb{R}^n$ to itself that preserves distances (*i.e.*, $d(\mathbf{x}, \mathbf{y}) = d\big(T(\mathbf{x}), T(\mathbf{y})\big)$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ — such maps also preserve angles).[9] More precisely, superposition phrases of the form, "Place figure A so that points B coincide with points C," may be interpreted as saying, "Find a rigid motion of $\mathbb{R}^n$ that maps points B to points C." Indeed, it seems that the inability of classical geometry to justify the notion of superposition resulted from the lack of a precise definition for rigid motions.

EXERCISES

1. Give a detailed verification of the assertion $T[\mathbf{x} + S] = T(\mathbf{x}) + T[S]$ which appears in Example 1 above.

2. If V and W are vector spaces over a field $\mathbb{F}$, a map $T : V \to W$ is called *affine* if it has the form $T(\mathbf{v}) = S(\mathbf{v}) + \mathbf{w}_0$, where S is linear and $\mathbf{w}_0 \in W$. Prove that

$$T\big(\, t\mathbf{x} \;+\; (1-t)\mathbf{y}\,\big) \;=\; tT(\mathbf{x}) \;+\; (1-t)T(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in V$ and $t \in \mathbb{F}$.

[8]See Exercise V.2.5 for a full explanation.

[9]The fact that rigid motions are geometrical symmetries follows because they all lie in the group $\mathrm{Aff}(V)$ defined previously. A proof is given in the Addendum to Appendix A.

3. Suppose that $\{\mathbf{v}_0, \cdots, \mathbf{v}_n\}$ and $\{\mathbf{w}_0, \cdots, \mathbf{w}_n\}$ are affine bases for the vector space V. Prove that there is a unique $T \in \mathrm{Aff}(V)$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for $i = 0, \cdots, n$. [*Hint:* The sets $\{\mathbf{v}_1 - \mathbf{v}_0, \cdots, \mathbf{v}_n - \mathbf{v}_0\}$ and $\{\mathbf{w}_1 - \mathbf{w}_0, \cdots, \mathbf{w}_n - \mathbf{w}_0\}$ are bases. Apply Theorem A.13(v).]

4. Suppose $\mathbb{F} \neq \mathbb{Z}_2$ is a field, V is a vector space over $\mathbb{F}$, and $T : V \to V$ is a 1–1 onto map satisfying the condition in Exercise 2:

$$T\big(t\mathbf{x} + (1-t)\mathbf{y}\big) = tT(\mathbf{x}) + (1-t)T(\mathbf{y})$$

Prove that T is an affine transformation. [*Hint:* Set $S(\mathbf{v}) = T(\mathbf{v}) - T(\mathbf{0})$ and prove S is linear. Observe that S maps $\mathbf{0}$ to itself. — If the problem seems too hard as stated, try to prove if for fields in which $1 + 1 \neq 0$.]

5. Let $f : (S, \Pi, d) \to (S', \Pi', d')$ be an isomorphism of geometrical incidence spaces, and let P and Q be nonempty geometrical subspaces of S in the extended sense (*i.e.*, we include the possibility that either consists of one point). If "$\star$" is the join construction described in the notes, prove that $f[P \star Q] = f[P] \star f[Q]$. [*Hint:* recall that the join of two geometrical subspaces is the unique smallest geometrical subspace containing both of them.]

6. Let $f : (S, \Pi, d) \to (S', \Pi', d')$ be an isomorphism of geometrical incidence spaces.

(i) Prove that (S, Π, d) is regular if and only if (S', Π', d') is.

(ii) Prove that (S, Π, d) is an affine n-space if and only if (S', Π', d') is.

(iii) Let $m > 2$ be an integer. Prove that every line in (S, Π, d) contains exactly m points if and only if the same is true for (S', Π', d'). Explain why a similar conclusion holds if we substitute "infinitely many" for "exactly m" in the statement.

7. Let V be a vector space over a field of characteristic zero (*i.e.*, no iterated finite sum $1 + \cdots + 1$ is equal to zero), and let $T \in \mathrm{Aff}(V)$ be an element such that some finite power $T^n = T \circ \cdots \circ T$ is the identity. Prove that there is some point $\mathbf{x} \in V$ such that $T(\mathbf{x}) = \mathbf{x}$ (*i.e.*, a **fixed point**). [*Hints:* Let $\mathbf{v} \in V$ be arbitrary, and consider the vector

$$\frac{1}{n} \sum_{i=0}^{n-1} T^i(\mathbf{v}) \ .$$

Note that the transformation T may have more than one fixed point; for example, if we take the reflection map T on $\mathbb{R}^2$ which sends (x, y) to $(x, -y)$, then T^2 is the identity and T sends every point on the x-axis to itself.

CHAPTER III

CONSTRUCTION OF PROJECTIVE SPACE

In this chapter we describe the classical process of adding points at infinity to an affine plane or a 3-dimensional affine space. The objects obtained in this manner are called **projective planes** or **projective spaces**, and predictably they are one of the main objects of attention in projective geometry.

1. Ideal points and lines

> Extending the space ... [is a] fruitful method for extracting understandable results from the bewildering chaos of special cases. — **J. Dieudonné (1906–1992)**

In calculus — particularly in the study of limits — it is frequently convenient to add one or two *numbers at infinity* to the real number system.[1] Among the reasons for this are the following:

(i) It allows one to formulate otherwise complicated notions more understandable (for example, infinite limits).

(ii) It emphasizes the similarities between the infinite limit concept and the ordinary limit concept.

(iii) It allows one to perform formal manipulations with limits much more easily.

For example, suppose we add a single point at infinity (called ∞ as usual) to the real numbers. If f is a real-valued rational function of the form $f(t) = p(t)/q(t)$, where p and q are polynomials with no common factors and q is not identically zero, then strictly speaking f is not definable at the roots of q. However, an inspection of the graph of f suggests defining its value at these points to be ∞, and if this is done the function is also continuous at the roots of q (in an appropriate sense). Proceeding further along these lines, one can even define $f(\infty)$ in such a way that f is continuous at ∞; the limit value may be a finite number or ∞, depending upon whether or not the degree of p is less than or equal to the degree of q (in which case the limit value is finite) or the degree of p is greater than the degree of q (in which case the limit value is infinite).

The discussion above illustrates the ideas presented in the following quotation from previously cited the book by R. Winger.[2]

---

[1]Of course, if this is done then one must also recognize that the *numbers at infinity* do not necessarily have all the useful properties of ordinary real numbers. The existence of such difficulties has been recognized since ancient times, and in particular this is implicit in the celebrated paradoxes which are attributed to Zeno of Elea ($c.$ 490 B.C.E. – $c.$ 425 B.C.E.).

[2]Winger, *Introduction to Projective Geometry*, pp. 31–32.

> Such exceptions are not uncommon in mathematics but they can frequently be avoided by aid of appropriate expedients. Often it suffices to modify definitions or merely adopt contentions of language. But sometimes new postulates or assumptions are required. This in algebra we might say that the quadratic equation $x^2 - 2ax + a^2 = 0$ has only one root a. For the sake of uniformity however it is customary to say that the equation has *two equal roots*. Here a change of language is all that is needed. On the other hand if the equation $x^2 + x + 1 = 0$ is to have any root it is necessary to extend the domain of numbers to include the imaginary numbers. With these conventions — that a repeated root counts for two and that imaginary roots are to be accepted equally with real — we can say *every* quadratic equation has two roots.
>
> Again we might say that a circle cuts a line of its plane in two points, one point or no point. But with the proper modifications we can make the geometry conform to the algebra. Thus a tangent is considered as meeting the curve in "two coincident points." But in order that the statement shall be true universally it is necessary to introduce a new class of points, the "imaginaries." Imaginary points correspond to the imaginary numbers of algebra [*i.e.*, their coordinates are given by complex numbers]. If in solving the equations of line and circle the roots turn out to be imaginary, the points of intersection are said to be imaginary. "No point" is now replaced by "two imaginary points" when without exception sl a line cuts a circle in two points — real and distinct, coincident or imaginary. The new statement not only serves every purpose of the old but is really more descriptive of the true relation of line and circle.

The preceding quotation illustrates that "imaginary numbers" in algebra have geometric analogs which can be viewed as "imaginary points." Similarly, **the previously discussed algebraic and analytic notions of "numbers at infinity" have geometric analogs which can be viewed as "points at infinity,"** and these play a fundamental role in projective geometry.

One intuitive motivation for considering such points at infinity arose in connection with the mathematical theory of *perspective drawing* which was developed in the $15^{\rm th}$ and $16^{\rm th}$ centuries. In modern and informal terms, the goal of this theory is to describe the photographic images of physical objects on a planar surface such as a projection screen, and artist's canvas, or a sheet of paper. Everyday experience with viewing photographs shows that some properties of objects are faithfully reflected by a photographic image while others are not. For example, distances often change drastically and two physical objects with the same measurements usually project to images with quite different measurements. On the other hand, the photographic images of a physical line will be contained in a line on the image plane.

We shall be particularly interested in a curious phenomenon involving parallel lines.[3] If one sees enough examples, it becomes apparent that the images of parallel lines are not necessarily parallel, and if the images are not parallel then the images of all lines parallel to the two given lines appear to meet at some point on the horizon; furthermore, all these horizon points lie appear to lie on a single line which defines the horizon (see Figure III.1[4] below). It is possible to give a mathematical explanation for these empirical observations, but we shall not do so here. A more detailed discussion appears in Section IV.2 of the online reference

`http://math.ucr.edu/~res/math133/geometrynotes4a.pdf`

that was cited in the Preface.

---

[3] This was known in ancient times and is mentioned in the work of Marcus VITRUVIUS Pollio (*c.* 80 B. C. E. – 25 B. C. E.) titled *De Architectura*.

[4] Source: `http://www.math.nus.edu.sg/aslaksen/projects/perspective/alberti.htm`

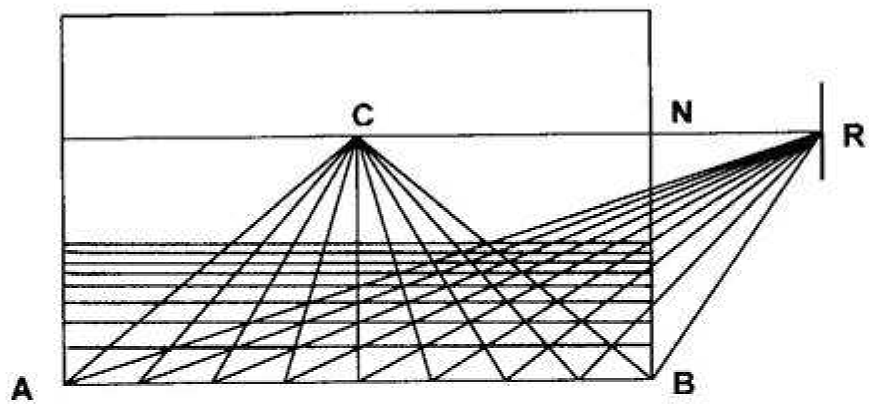


Figure III.1

Such pictures lead to speculation whether we should think of lines as having *points at infinity*, such that every line has exactly one such point, two parallel lines have the same point at infinity, and the points at infinity for lines on a given plane all lie on a *line at infinity*. These ideas emerged near the beginning of the 17^{th} century, and in particular they were developed and extended in the writings of G. Desargues (1591–1661) and J. Kepler (1571–1630).

The main purpose of this section is to provide a mathematical setting for Desargues' and Kepler's intuitive ideas. In particular, the following continuation of the previous quotation contains the main motivation for the construction of projective space (and much of projective geometry):

> To say that two parallel lines do not meet is like saying that certain lines have no point of intersection with a circle. There we found that the exception could be removed by introducing imaginary intersections. In an exactly analogous fashion we may introduce a second new class of points into geometry, *points at infinity*, which will serve for the "intersections of parallel lines."

The formal process for adding points at infinity to the Euclidean plane is best described as follows:

Definition. Let $(P, \mathcal{L})$ be an affine plane, and let L be a line in P. The L-**direction** in P (or the **direction** of P) consists of L and all lines parallel to L.

Note. We are using the notation for incidence plane described in Section II.1 rather than the more general notation in Section II.5. The translation is straightforward: If $(P, \mathcal{L})$ is one description and (P, Π, d) is the other, then $d(P) = 2$, $\Pi = \mathcal{L} \cup \{P\}$ and $d = 1$ on $\mathcal{L}$.

Theorem III.1. *Two directions in P are either disjoint or identical.*

Proof. Consider the binary relation $\sim$ on lines defined by $L \sim M$ if and only if $L = M$ or $L||M$. By Theorem II.7 this is an equivalence relation. The L-direction is merely the equivalence class of L with respect to this relation and will be denoted by $[L]$.■

Definition. If P is an affine plane, the *projective extension* pf P, denoted by $P^\wedge$, consists of P together with all directions in P. An element of $P^\wedge$ is an ordinary point if it is a point in P

and an *ideal point* if it is a direction in P. It follows that a point in is either ordinary or ideal but not both.[5]

If L is a line in P, then its *projective extension* to $P^\wedge$ consists of all points of L together with the direction $[L]$, and it is denoted by $L^\wedge$. The *ideal line* in $P^\wedge$ consists of all ideal points in $P^\wedge$ and is denoted by L_∞ (the line of points at infinity or the *line at infinity*).

EXAMPLES INVOLVING USES OF POINTS AT INFINITY. The most immediate reason for introducing points at infinity is that they can be used to simplify some of the statements which appeared in earlier chapters. For example, the statement

> The lines L, M and N are either concurrent (pass through a single point) or are parallel in pairs.

in the conclusion of Theorem II.25 translates to

> The projective extensions of the lines L, M and N contain a common point (*i.e.*, the extended lines $L^\wedge$, $M^\wedge$ and $N^\wedge$ are concurrent).

Furthermore, the statement

> There is a point C which lies on line AB or else there are two lines L, and M such that $L||AB$ and $M||AB$,

which is the conclusions of Theorem II.28 and II.29, translates to

> There are three lines AB, L and M such that either (i) L and M meet at a point $C \in AB$, or else the lines L and M are both parallel to and their common point at infinity lies on the projective extension of AB (*i.e.*, if the projective extensions of L and M meet at the point C, then C lies on the projective extension of the line AB).

Since the hypotheses of Theorems II.27–29 are very similar (the only difference being parallelism assumptions about these lines) the second translated statement illustrates a corresponding similarity in the conclusions that one might suspect (or at least hope for). In fact, the conclusion of Theorem II.27 (that one has three pairs of parallel lines) fits into the same general pattern, for in this case the conclusion reduces to the collinearity of the directions $A^\wedge$, $B^\wedge$ and $C^\wedge$, where these directions contain the pairs $\{L, L'\}$, $\{M, M'\}$, and $\{N, N'\}$ respectively.

One conceivable objection to *ideal points* or *points at infinity* is the impossibility of visualizing such entities. The mathematical answer to such objection is contained in the following quotation from an article by O. Veblen (1880–1960):[6]

> Ordinary points are just as much idealized as are the points at infinity. No one has ever seen an actual point [with no physical width, length or thickness] or realized it by an experiment of any sort. Like the point at infinity it is an ideal creation which is useful for some of the purposes of science.

[5] A rigorous proof of this fact requires some technical propositions from set theory. The details of this justification are not important for the rest of these notes, but for the sake of completeness here is the proof: Suppose that $\mathbf{x} = [L]$ for suitable $\mathbf{x}$ and L. Then there is a line M such that $\mathbf{x} \in \mathbf{M}$ and L is parallel to M or equal to M. This would imply that $\mathbf{x} \in M$ and $[M] = [L] = \mathbf{x}$, so we would have objects a and b such that $a \in b$ and $b \in a$; the standard mathematical foundations for set theory contain an assumption which implies that such situations never arise. For further information, see Proposition 4.2 in the following online notes: `http://math.ucr.edu/~res/math144/setsnotes3.pdf`

[6] Encyclopædia Britannica, 14$^{\text{th}}$ Edition (1956),Vol. 18, p.173 (article on Projective Geometry)

Here is a slightly different response:[7]

> [With the introduction of points at infinity,] The propositions of projective geometry acquire a simplicity and a generality that they could not otherwise have. Moreover, the elements at infinity give to projective geometry a degree of unification that greatly facilitates the thinking in this domain and offers a suggestive imagery that is very helpful ... On the other hand, projective geometry stands ready to abandon these ... whenever that seems desirable, and to express the corresponding propositions in terms of direction of a line ... to the great benefit of ... geometry. ...
>
> The extra point which projective geometry claims to add to the Euclidean [or affine] line is [merely] the way in which projective geometry accounts for the property of a straight line which Euclidean [or affine] geometry recognizes as the "direction" of the line. The difference between the Euclidean [or affine] line and the projective line is purely verbal. The geometric content is the same. ... Such a change in nomenclature does not constitute an [actual] increase in the geometric content.

To summarize the preceding quotations, sometimes it is convenient to work in a setting where one has points at infinity, and in other cases it is more convenient to work in a setting which does not include such objects. This is very closely reflects the standard usage for numbers at infinity. They are introduced when they are useful — with proper attention paid to the differences between them and ordinary numbers — and not introduced when it is more convenient to work without them.

EXERCISES

In the exercises below, assume that P is an affine plane and that $x^\wedge$ (where $x = P$ or a line in P) is defined as in the notes.

1. Prove that every pair of lines in $P^\wedge$ (as defined above) has a common point in $P^\wedge$. [*Hint:* There are three cases, depending upon whether one has extensions of two ordinary lines that have a common point in P, extensions of two ordinary lines that are parallel in P, or the extension of one line together with the line at infinity.]

2. Suppose that L_1, L_2, L_3, M_1, M_2, M_3 are six distinct lines in an affine plane P. Write out explicitly what it means (in affine terms) for the three points determined by $L_i^\wedge \cap M_i^\wedge$ (where $i = 1, 2, 3$) to be collinear in $P^\wedge$. Assume that the three intersection points are distinct.

3. Suppose that $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$ are six distinct points in an affine plane P. Write out explicitly what it means (in affine terms) for the three extended lines $(\mathbf{x}_i\mathbf{y}_i)^\wedge$ (where $i = 1, 2, 3$) to be concurrent in $P^\wedge$. Assume that the three ordinary lines $\mathbf{x}_i\mathbf{y}_i$ are distinct.

4. Prove that every line in $P^\wedge$ contains at least three points. There are two cases, depending upon whether the line is the extension of an ordinary line or the line at infinity, and the latter requires more work than the former.

[7] N. Altshiller Court (1881–1968), *Mathematics in Fun and in Earnest*, pp. 110, 112

2. Homogeneous coordinates

If $\mathbb{F}$ is a field and $\mathbb{F}^2$ is the coordinate affine plane described in Chapter II, then the construction in the previous section can of course be applied to $\mathbb{F}^2$. The purpose of this section is to introduce coordinates for the projective extension $(\mathbb{F}^2)^\wedge$ that are consistent with the usual coordinates for ordinary points. Since every point in $\mathbb{F}^2$ is specified by two scalars (the first and second coordinates), we shall not be able to describe a point entirely in terms of two coordinates. Instead, it will be necessary to use **three** coordinates to specify a point in $(\mathbb{F}^2)^\wedge$, with the understanding that **different triples of scalars may represent the same point.** This can be inconvenient sometimes, but it is an issue that already arises in elementary analytic geometry; specifically, when we try to specify a point in the ordinary plane $\mathbb{R}^2$ by polar coordinates, it is always necessary to remember that (r, θ), $(-r, \theta + \pi)$ and $(r, \theta + 2\pi)$ represent the same point in $\mathbb{R}^2$ if $r \neq 0$ (and of course all pairs of the form $(0, theta)$ represent the origin!).

To be more specific about the meaning of compatibility, we would like our coordinates for $(\mathbb{F}^2)^\wedge$ to have the following properties:

(i) The ordinary coordinates for a point $(x, y) \in \mathbb{F}^2$ should be easily retrievable from the systems of coordinates we shall introduce on $(\mathbb{F}^2)^\wedge$, and vice versa.

(ii) If we are given a line L in $\mathbb{F}^2$ defined by an equation of the form $ax + by + c = 0$ where a and b are not both zero, the coefficients a and b should be easily retrievable from the coordinates of the associated point at infinity $L^\wedge$, and vice versa.

Of course, in the second part the coefficients $a,\ b\ , c$ are not uniquely defined, and any nonzero multiple of these equations yields an equivalent equation for the line. This sort of ambiguity up to multiplication by some common nonzero factor is the key idea behind the definition of **homogeneous coordinates** for points in $(\mathbb{F}^2)^\wedge$.

Suppose first that L is a line through $\mathbf{0}$ in $\mathbb{F}^2$. Then points on this line have the form $(tx_1,\, tx_2)$, where $\mathbf{x} = (x_1,\, x_2)$ is a fixed nonzero vector. With some imagination, one might speculate about trying to define the coordinates for the point at infinity on L by something like $(\infty x_1,\, \infty x_2)$ or equivalently by

$$\left(\frac{x_1}{0}, \frac{x_2}{0} \right) .$$

Of course, we cannot naïvely do this in a logically sound manner (for example, if $\mathbb{F} = \mathbb{R}$ then $1/0 \ = \ 2/0 = \infty$ and $0/0$ is indeterminate), but we can express the concept using an ordered triple

$$(v_1,\, v_2,\, 0)$$

which is meant to suggest that we would divide the first two coordinates by zero if this made sense. As already noted, if we use such notation then we must also be ready to agree that every ordered triple of the form $(tv_1,\, tv_2,\, 0)$, where $t \neq 0$, is also a valid description fot the original line L. We can formalize this by saying that every such triple is a set of *homogeneous coordinates* for the point at infinity $L^\wedge$.

For the sake of uniformity, we would also like to describe coordinates for ordinary points as ordered triples $(y_1,\, y_2,\, y_3)$ such that if $t \neq 0$ and $(y_1,\, y_2,\, y_3)$ is a valid set of coordinates for a point, then so is $(ty_1,\, ty_2,\, ty_3)$. We do this by agreeing that if $\mathbf{x} \in \mathbb{F}^2$, then *every ordered triple of the form* $(tx_1,\, tx_2,\, t)$, *where* $t \neq 0$, *is a set of* homogeneous coordinates *for* $\mathbf{x}$.

The next result states that the preceding definitions of homogeneous coordinates for points of $(\mathbb{F}^2)^\wedge$ have the desired compatibility properties and that every ordered triple $(y_1, y_2, y_3) \neq (0,0,0)$ is a valid set of coordinates for some point in $(\mathbb{F}^2)^\wedge$.

THEOREM III.2. *Every nonzero element of $\mathbb{F}^3$ is a set of homogeneous coordinates for some point in $(\mathbb{F}^2)^\wedge$. Two nonzero elements are homogeneous coordinates for the same point if and only if each is a nonzero multiple of the other.*

Proof. Let $(y_1, y_2, y_3) \neq (0,0,0)$ in $\mathbb{F}^3$. If $y_3 \neq 0$ then (y_1, y_2, y_3) is a set of homogeneous coordinates for the ordinary point

$$\left(\frac{y_1}{y_3}, \frac{y_2}{y_3}\right) .$$

On the other hand, if $y_3 = 0$ then either $y_1 \neq 0$ or $y_2 \neq 0$ and the point has the form $(y_1, y_2, 0)$; the latter is a set of homogeneous coordinates for the point at infinity on the line joining the two distinct ordinary points $\mathbf{0}$ and (y_1, y_2); these points are distinct because at least one of the y_i is nonzero.

By definition, if $t \neq 0$ then (y_1, y_2, y_3) and (ty_1, ty_2, ty_3) are sets of homogeneous coordinates for the same point, so it is only necessary to prove the converse statement. Therefore suppose that (x_1, x_2, x_3) and (y_1, y_2, y_3) are homogeneous coordinates which represent the same point. There are two cases, depending upon whether or not this point is an ordinary point or an ideal point.

Suppose the point under consideration is the ordinary point $(z_1, z_2) \in \mathbb{F}^2$. Then there exist nonzero constants a and b such that

$$(x_1, x_2, x_3) \;=\; (az_1, , az_2, a) \qquad\qquad (y_1, y_2, y_3) \;=\; (bz_1, , bz_2, b) \;.$$

It follows immediately that $(y_1, y_2, y_3) \;=\; ba^{-1}\,(x_1, x_2, x_3)$.

Suppose now that the point under consideration is an ideal point, and choose $\mathbf{v} \in \mathbb{F}^2$ so that $\mathbf{v} \neq \mathbf{0}$ and the ideal point is contained in the line $\mathbf{0v}$ (the existence of such a line is guaranteed by the Euclidean Parallelism Property, which holds in $\mathbb{F}^2$). Let $\mathbf{v} = (v_1, v_2)$, so that the ideal point has homogeneous coordinates $(v_1, v_2, 0)$. In this case there exist nonzero constants a and b such that

$$(x_1, x_2, x_3) \;=\; (av_1, , av_2, 0) \qquad\qquad (y_1, y_2, y_3) \;=\; (bv_1, , bv_2, 0) \;.$$

It follows immediately that $y_3 = x_3 = 0$ and hence the vectors $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ satisfy $\mathbf{x} = a\mathbf{v}$ and $\mathbf{y} = b\mathbf{v}$; the latter implies that $\mathbf{y} = ba^{-1}\mathbf{x}$ and hence again in this case we conclude that (x_1, x_2, x_3) and (y_1, y_2, y_3) are nonzero multiples of each other.■

THEOREM III.3. *There is a $1-1$ correspondence between the points of the projective extension $(\mathbb{F}^2)^\wedge$ and the 1-dimensional vector subspaces of $\mathbb{F}^3$ such that a point $\mathbf{x}$ in the former corresponds to a 1-dimensional subspace V if and only if the set of all possible homogeneous coordinates for $\mathbf{x}$ is the set of nonzero vectors in V.*

Proof. Given $\mathbf{x} \in (\mathbb{F}^2)^\wedge$, let $V(\mathbf{x})$ denote the vector space spanned by all the homogeneous coordinates for $\mathbf{x}$. By the preceding result, we know that $V(\mathbf{x})$ is a 1-dimensional vector subspace because it contains a nonzero vector and every pair of nonzero vectors in $V(\mathbf{x})$ are nonzero scalar multiples of each other. Theorem 2 also implies that the correspondence sending $\mathbf{x}$ to $V(\mathbf{x})$ is onto (every 1-dimensional vector space is spanned by some nonzero vector). Thus it remains to show that the correspondence is 1–1.

Suppose that $\mathbf{x}$ and $\mathbf{y}$ are such that $V(\mathbf{x}) = V(\mathbf{y})$. There are three possible cases, depending upon whether both points are ordinary points, both points are ideal points, or one point is ordinary and the other is ideal.

CASE 1. Suppose both points are ordinary. Let $\mathbf{v} = (v_1, v_2)$ where $\mathbf{v} = \mathbf{x}$ or $\mathbf{y}$. If $V(\mathbf{x}) = V(\mathbf{y})$, then we know that $(x_1, x_2, 1)$ and $(y_1, y_2, 1)$ span the same 1-dimensional subspace and hence are nonzero multiples of each other. However, if c is a nonzero scalar such that $(x_1, x_2, 1) = c\,(y_1, y_2, 1)$, then it follows immediately that $c = 1$, which implies that $bfx = \mathbf{y}$.

CASE 2. Suppose both points are ideal. Suppose that the points in question are the ideal points on the lines joining the origin in $\mathbb{F}^2$ to the nonzero points $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. It then follows that the vectors $(u_1, u_2, 0)$ and $(v_1, v_2, 0)$ span the same 1-dimensional subspace and hence are nonzero multiples of each other. But this implies that $\mathbf{u}$ and $\mathbf{v}$ are also nonzero multiples of each other, which in turn means that the lines joining $\mathbf{0}$ to $\mathbf{u}$ and $\mathbf{v}$ are equal and hence have the same ideal point. Therefore the map taking $\mathbf{x}$ to $V(\mathbf{x})$ is 1–1 on the set of ideal points.

CASE 3. Suppose one point is ordinary and the other is ideal. If the ordinary point has homogeneous coordinates given by $(x_1, x_2, 1)$ and the ideal point has homogeneous coordinates given by $(y_1, y_2, 0)$, then once again these two vectors must be nonzero multiples of each other. Since the third coordinate of the first vector is equal to 1, this is impossible, and thus we see that ordinary points and ideal points cannot determine the same 1-dimensional subspace of $\mathbb{F}^3$.■

EXERCISES

1. Suppose that the line in $\mathbb{F}^2$ is defined by the equation

$$ax + by = c$$

where not both of a and b are zero. Show that homogeneous coordinates for the point $L^\wedge$ are given by $(-b, a, 0)$. [*Hint:* What is the equation of the line through $(0,0)$ which is parallel to L?]

2. Suppose that $\mathbb{F} = \mathbb{R}$ (the real numbers), and $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ is nonzero. Let (y_1, y_2, y_3) be a set of homogeneous coordinates for $\mathbf{x}$ such that $y_1^2 + y_2^2 = 1$. Prove that $|y_3|$ is the reciprocal of the distance from $\mathbf{x}$ to the origin. [*Hint:* Start by explaining why $x_1^2 + x_2^2 > 0$.]

> *Note.* Exercise 2 reflects one reason why ideal points are also known as *points at infinity*. Their last coordinates always vanish, so a formal candidate for their difference to the origin would be the reciprocal of 0, which we often think of as ∞, at least informally.

3. Equations of lines

In Theorem I.7 we noted that lines in $\mathbb{F}^2$ are precisely the subsets defined by linear equations of the form $ax + by + c = 0$, where a and b are not both zero. An analogous characterization for lines in $(\mathbb{F}^2)^\wedge$ will be given in this section.

THEOREM III.4. *Let L be a line in $\mathbb{F}^2$ defined by the equation $ax + by + c = 0$ as above, and let $L^\wedge$ be the extension of L to a line in $(\mathbb{F}^2)^\wedge$. Then $L^\wedge$ consists of all points $\mathbf{x}$ which are representable by homogeneous coordinates (u_1, u_2, u_3) satisfying*

$$a\,u_1 + b\,u_2 + c\,u_3 = 0 \; .$$

Remark. If one set of homogeneous coordinates for a point $\mathbf{x}$ satisfies the equation above, then so does every other set, for every other set has the form (ku_1, ku_2, ku_3) for some $k \neq 0$ and hence $a\,u_1 + b\,u_2 + c\,u_3 = 0$ implies

$$a(ku_1) + b(ku_2) + c(ku_3) = k(a\,u_1 + b\,u_2 + c\,u_3) = k \cdot 0 = 0 \; .$$

Proof. If (x, y) is an ordinary point if $(\mathbb{F}^2)^\wedge$, then it belongs to L and has homogeneous coordinates $(x, y, 1)$. Therefore every ordinary point of L has homogeneous coordinates which satisfy the homogeneous linear equation in three variables that is displayed in the theorem. Furthermore, by Exercise 1 for the preceding section, the ideal point of $(\mathbb{F}^2)^\wedge$ has homogeneous coordinates given by $(-b, a, 0)$, and hence the ideal point also has homogeneous coordinates which satisfy the homogeneous linear equation $a\,u_1 + b\,u_2 + c\,u_3 = 0.\Box$

Conversely, suppose that (x, y) is an ordinary point not on the line L. Then (x, y) has homogeneous coordinates $(x, y, 1)$, and these coordinates do not satisfy the equation $a\,u_1 + b\,u_2 + c\,u_3 = 0$. By the remark preceding the proof, it follows that an ordinary point lies on L (equivalently, $(\mathbb{F}^2)^\wedge$) if and only if it has homogeneous coordinates which satisfy the equation $a\,u_1 + b\,u_2 + c\,u_3 = 0$.

Finally suppose $\mathbf{z}$ is an ideal point whose homogeneous coordinates satisfy $a\,z_1 + b\,z_2 + c\,z_3 = 0$. Since $\mathbf{z}$ is an ideal point, we also know that $z_3 = 0$. Therefore we must have $a\,z_1 + b\,z_2 = 0$. However, the solution space for this nontrivial equation is 1-dimensional by Theorem A.10, and hence the nonzero vector (z_1, z_2) must be a nonzero multiple of $(-b, a)$. Therefore Exercise 1 of the preceding section implies that $\mathbf{z}$ must be the ideal point of $(\mathbb{F}^2)^\wedge$, and this completes the proof of the theorem.■

There is a similar characterization of the line at infinity.

THEOREM III.5. *A point of $(\mathbb{F}^2)^\wedge$ is an ideal point if and only if it has homogeneous coordinates which satisfy the equation $u_3 = 0$.*

Proof. Suppose that we are given an ideal point. By construction it has homogeneous coordinates which satisfy $u_3 = 0$. Conversely, if we are given an ordinary point $\mathbf{x}$, then as noted before we know that $\mathbf{x}$ has homogeneous coordinates of the form $(x_1, x_2, 1)$, and hence **every** set of homogeneous coordinates (u_1, u_2, u_3) must satisfy $u_3 \neq 0$. Thus the ideal points are characterized by the vanishing of the third homogeneous coordinate.■

The previous two theorems yield the desired characterization of lines in $(\mathbb{F}^2)^\wedge$.

THEOREM III.6. *A set of points X in $(\mathbb{F}^2)^\wedge$ is line if and only if there exist $a,\ b,\ c \in \mathbb{F}$ not all zero such that $\mathbf{y} \in L$ if and only if $\mathbf{y}$ has homogeneous coordinates $(y_1,\, y_2,\, y_3)$ satisfying $a\,y_1 \;+\; b\,y_2 \;+\; c\,y_3 \;=\; 0$.*

Proof. By the preceding two results, every line in $(\mathbb{F}^2)^\wedge$ is defined by a nontrivial linear homogeneous equation of the type described. Conversely, suppose that $(a, b, c) \neq (0, 0, 0)$. If we have $a = b = 0$, then the linear homogeneous equation is equivalent to the equation $y_3 = 0$, which defines the line at infinity. On the other hand, if at least one of $\{a,\, b\}$ is nonzero, then by Theorem 4 we know that the equation defines the extension $L^\wedge$ of the line L in $\mathbb{F}^2$ with equation $ax + by + c = 0$.■

EXERCISES

1. Find the equations of the lines joining points in $(\mathbb{R}^2)^\wedge$ with the following homogeneous coordinates:

(i) $(1, 3, 0)$ and $(0, 5, -1)$

(ii) $(2, 5, -3)$ and $(3, -2, 0)$

(iii) $(2, 1, 0)$ and $(-1, 1, 0)$

(iv) $(4, -6, 3)$ and $(4, -6, 1)$

2. Find the homogeneous coordinates of the intersection points of the following lines in $(\mathbb{R}^2)^\wedge$:

(i) $x_1 \;+\; x_3 \;=\; 0$ and $2x_1 \;+\; 3x_2 \;+\; x_3 \;=\; 0.$

(ii) $2x_1 \;+\; 3x_2 \;-\; 4x_3 \;=\; 0$ and $x_1 \;-\; x_2 \;+\; 3x_3 \;=\; 0.$

(iii) $2x_1 \;+\; x_2 \;=\; 0$ and $-x_1 \;+\; x_2 \;=\; 0.$

(iv) $4x_1 \;-\; 6x_2 \;+\; 3x_3 \;=\; 0$ and $4x_1 \;-\; 6x_2 \;+\; x_3 \;=\; 0.$

4. Higher-dimensional generalizations

In Chapter II we generalized the geometrical properties of coordinate affine planes and 3-spaces to arbitrary dimensions. The purpose of this section is to show that preceding construction for adding ideal points to $\mathbb{F}^2$ can be extended to the affine spaces associated to arbitrary n-dimensional vector spaces over a given field $\mathbb{F}$, where n is an arbitrary positive integer.

NOTATION. The projective extension $(\mathbb{F}^2)^\wedge$ that was defined and studied in the preceding three sections will be called the *coordinate projective plane over* $\mathbb{F}$ and will be denoted by the symbol $\mathbb{FP}^2$; this conforms to standard mathematical terminology.

The results of the preceding two sections imply the following two basic properties of $\mathbb{FP}^2$:

(1) The *points* of $\mathbb{FP}^2$ correspond to *one-dimensional* vector subspaces of $\mathbb{F}^3$.

(2) The *points* of $\mathbb{FP}^2$ correspond to *two-dimensional* vector subspaces of $\mathbb{F}^3$, and a point $\mathbf{x}$ lies on a line L if and only if the one-dimensional subspace V associated to $\mathbf{x}$ is contained in the two-dimensional subspace W associated to L.

The first of these is just a restatement of Theorem 3. The second statement follows because a subset of $\mathbb{F}^3$ is a 2-dimensional vector subspace if and only if it is the set of solutions to a nontrivial linear homogeneous equation in three variables; this is essentially a special case of the characterization of $)n-1)$-dimensional vector subspaces of an n-dimensional vector space in the exercises for Section 4 of the Appendix.

Motivated by this identification, if V is a finite-dimensional left or right[8] vector space over a skew-field $\mathbb{F}$, we shall define the *projective space with coefficients in a* **nonzero** *vector space* V as follows: Its *points* are the elements of the set $\mathcal{S}_1(V)$ of 1-dimensional vector subspaces of V. Note that if W is a nonzero vector subspace of V, then $\mathcal{S}_1(W)$ is contained in $\mathcal{S}_1(V)$. The *geometric subspaces* of $\mathcal{S}_1(V)$ are given by all subsets of the form $\mathcal{S}_1(W)$, where W is a vector subspace of dimension ≥ 2 and the geometric *dimension* of $\mathcal{S}_1(W)$ is equal to $\dim W - 1$. — The shift of dimensions is consistent with our previous construction of homogeneous coordinates for $\mathbb{FP}^2$; in particular, if $\dim V = n$, then the dimension of $\mathcal{S}_1(V)$ is equal to $n-1$.

If we adopt the conventions of Section II.5 for geometrical subspaces of dimension 0 and -1 (one point subsets are zero-dimensional and the empty set is (-1)-dimensional, then the equation

$$d\,(\,\mathcal{S}_1(W)\,) \;=\; \dim W \;-\; 1$$

also holds in these extended cases because $\dim X = 1$ implies $cS_1(X) = \{\, X \,\}$ and also

$$cS_1\,(\,\{\mathbf{0}\}\,) \;=\; \varnothing$$

(since the zero subspace has no 1-dimensional subspaces).

NOTATION. If $\mathbf{x} \in \mathcal{S}_1(V)$, then symbols like $\widetilde{\mathbf{x}}$ will denote nonzero vectors in V which belong to the 1-dimensional subspace $\mathbf{x}$, and such a vector $\widetilde{\mathbf{x}}$ will be called a set of *homogeneous coordinates* for $\mathbf{x}$.

If $V = \mathbb{F}^{n+1}$, then $\mathcal{S}_1(V)$ will be callled the *standard (coordinate) projective n-space over* $\mathbb{F}$ and it is usually denoted by $\mathbb{FP}^n$. As with affine spaces, the projective spaces whose geometrical

[8] The geometric significance of doing everything for both left and right vector spaces will be apparent in Chapter V, but if one is only interested in cases where $\mathbb{F}$ is a field then the distinction is unnecessary.

usefulness is most evident are the projective spaces over the real numbers $\mathbb{R}$; these are the so-called *real projective spaces*, and in some cases they are also simply called *projective spaces of n dimensions*. The corresponding objects over the complex numbers $\mathbb{C}$ (the *complex projective spaces*) are nearly as important; the Fundamental Theorem of Algebra suggests one reason for this (however, there are also others). Other types of projective spaces are useful in various contexts which are beyond the scope of these notes.

THEOREM III.7. *If V is an n-dimensional vector space (where $n \geq 3$), then $\mathcal{S}_1(V)$ is a regular $(n-1)$-dimensional geometrical incidence space with respect to the notions of geometrical subspace and dimension that are defined above.*

Proof. We shall first verify the conditions **(G-1)**–**(G-3)** in order, and in the next theorem we shall prove a strengthened version of **(G-4)** in the next theorem.

PROOF OF **(G-1).** Let $X_0, \cdots, X_k$ be 1-dimensional subspaces of V that are not contained in any subspace $\mathcal{S}_1(W)$ of dimension less than k; *i.e.*, there is no vector subspace $W \subset V$ such that $\dim W \leq k$ and $X_i \subset W$ for all i. By the sum formula for dimensions of vector subspaces, we know that $\dim(X_0 + \cdots + X_k) \leq k+1$; furthermore, if strict inequality holds, then $X_0 + \cdots + X_k$ is a vector space of dimension $\leq k$ containing each X_i, and hence we know that all the subspaces X_i belong to some geometric subspace $\mathcal{S}_1(W)$ of dimension less than k. Since we are assuming this does not happen, it follows that $\dim(X_0 + \cdots + X_k) = k+1$, and accordingly the k-plane

$$\mathcal{S}_1(X_0 + \cdots + X_k)$$

contains all the 1-dimensional subspaces X_i. To prove the uniqueness part of **(G-1)**, suppose that $\mathcal{S}_1(W)$ is a k-plane such that $X_i \subset W$ for all i. It follows immediately that $X_0 + \cdots + X_k \subset W$ and

$$k+1 \;=\; \dim(X_0 + \cdots + X_k) \;\leq\; \dim W \;=\; k+1$$

and hence $(X_0 + \cdots + X_k) = W$ by Theorem A.8 so that $\mathcal{S}_1(X_0 + \cdots + X_k)$ is the unique k-plane containing all the X_i.■

PROOF OF **(G-2).** Suppose that $X_i \in \mathcal{S}_1(W)$ for $0 \leq i \leq m$, and assume that the set $\{X_0, \cdots, X_m\}$ is independent. By the previous proof, the unique m-plane containing the X_i is $\mathcal{S}_1(X_0 + \cdots + X_m)$; since $X_i \subset W$ by our hypotheses, it follows that $(X_0 + \cdots + X_m) \subset W$, and therefore we also have $\mathcal{S}_1(X_0 + \cdots + X_m) \subset \mathcal{S}_1(W)$.■

PROOF OF **(G-3).** Suppose that $\mathcal{S}_1(W)$ is a k-plane, so that $\dim W = k+1$. Let $\mathbf{w}_0, \cdots, \mathbf{w}_k$ be a basis for W, and let for each i such that $0 \leq i \leq k$ let X_i be the 1-dimensional vector subspace spanned by $\mathbf{w}_i$. Then $\{X_0, \cdots, X_k\}$ is a set of $k+1$ distinct points in $\mathcal{S}_1(W)$.■

The next result will show that a strengthened form of **(G-4)** holds for $\mathcal{S}_1(V)$.

THEOREM III.8. *If P and Q are geometrical subspaces of $\mathcal{S}_1(V)$ then*

$$d(P) \;+\; d(Q) \;=\; d(P \star Q) \;+\; d(P \cap Q)\ .$$

The difference between this statement and **(G-4)** is that the latter assumes $P \cap Q \neq \varnothing$, but the theorem contains no such assumption.

Proof. We shall first derive the following equations:

(i) $\mathcal{S}_1(W_1 \cap W_2) = \mathcal{S}_1(W_1) \cap \mathcal{S}_1(W_2)$

(ii) $\mathcal{S}_1(W_1 + W_2) = \mathcal{S}_1(W_1) \star \mathcal{S}_1(W_2)$

Derivation of (i). X is a 1-dimensional subspace of $W_1 \cap W_2$ if and only if X is a 1-dimensional subspace of both W_1 and W_2.

Derivation of (ii). If W is a vector subspace of U, then $\mathcal{S}_1(W) \subset \mathcal{S}_1(U)$. therefore $\mathcal{S}_1(W_i) \subset \mathcal{S}_1(W_1 + W_2)$ for $i = 1, 2$. Therefore we also have $\mathcal{S}_1(W_1 + W_2) \subset \mathcal{S}_1(W_1) \star \mathcal{S}_1(W_2)$. To prove the reverse inclusion, choose U such that $\mathcal{S}_1(U) = \mathcal{S}_1(W_1) \star \mathcal{S}_1(W_2)$. Then $W_i \subset U$ for $I = 1, 2$ follows immediately, so that $W_1 + W - 2 \subset U$. Consequently we have

$$\mathcal{S}_1(W_1 + W_2) \subset \mathcal{S}_1(U) \subset \mathcal{S}_1(W_1 + W_2)$$

which immediately yields $\mathcal{S}_1(W_1 + W_2) = \mathcal{S}_1(U) = \mathcal{S}_1(W_1 + W_2)$.

To prove the theorem, note that

$$\begin{gathered} d\,(\,\mathcal{S}_1(W_1) \star \mathcal{S}_1(W_2)\,) = d\,(\,\mathcal{S}_1(W_1 + W_2)\,) = \\ \dim(W_1 + W_2) - 1 = \dim W_1 + \dim W_2 - \dim W_1 \cap W_2 - 1 = \\ (\dim W_1 - 1) + (\dim W_2 - 1) - (\dim(W_1 \cap W_2) - 1) = \\ d\,(\,\mathcal{S}_1(W_1)\,) + d\,(\,\mathcal{S}_1(W_2)\,) - d\,(\,\mathcal{S}_1(W_1 \cap W_2)\,)\ .\blacksquare \end{gathered}$$

EXAMPLE 1. If $\mathcal{S}_1(V)$ is 2-dimensional, then Theorem 9 states that every two lines in $\mathcal{S}_1(V)$ have a common point because $d(L_1 \star L_2) \le 2 = d\,(\,\mathcal{S}_1(V)\,)$ implies

$$d(L_1 \cap L_2) = d(L_1) + d(L_2) - d(L_1 \star L_2) \ge 1 + 1 - 2 = 0\ .$$

EXAMPLE 2. Similarly, if $\mathcal{S}_1(V)$ is 3-dimensional, then every pair of planes has a line in common because $d(P_1 \star P_2) \le 3 = d\,(\,\mathcal{S}_1(V)\,)$ implies

$$d(P_1 \cap P_2) = d(P_1) + d(P_2) - d(P_1 \star P_2) \ge 2 + 2 - 3 = 1\ .$$

The next result will play a significant role in Chapter IV.

THEOREM III.9. *If V as above is at least 2-dimensional, then every line in $\mathcal{S}_1(V)$ contains at least three points.*

Proof. Let W be a 2-dimensional subspace of V, and let $\mathbf{w}_1$ and $\mathbf{w}_2$ form a basis for W. Then $\mathcal{S}_1(W)$ contains the three distinct points $\mathsf{Span}(\mathbf{w}_1)$, $\mathsf{Span}(\mathbf{w}_2)$, and $\mathsf{Span}(\mathbf{w}_1 + \mathbf{w}_2)$.$\blacksquare$

Projective extension of the affine space V

Given an n-dimensional vector space V over a field $\mathbb{F}$, we shall now construct a projective extension of the affine space structure on V which generalizes the previous construction of $\mathbb{FP}^2$ from $^{\mathbf{2}}$. The central object in this construction is a 1–1 mapping J_V from V to $\mathbb{P}(V) = \mathcal{S}_1(V \times \mathbb{F})$ which sends $\mathbf{v} \in V$ to the 1-dimensional vector subspace of $V \times \mathbb{F}$ spanned by $(\mathbf{v},\, 1)$. Notice that $V \times \mathbb{F}$ is a vector space with addition and scalar multiplication defined coordinatewise. As in the previous construction, a point is said to be *ordinary* if it lies in the image of J_V, and the remaining points, which are represented by homogeneous coordinates of the form $(\mathbf{v}, 0)$, where $\mathbf{v} \neq \mathbf{0}$, are called *ideal points*. Once again, the ideal point with homogeneous coordinates $(\mathbf{v}, 0)$

is the point at infinity on the line joining $\mathbf{0}$ to $\mathbf{v}$. We may summarize the preceding discussion as follows:

THEOREM III.10. *A point in* $\mathbb{P}(V) = \mathcal{S}_1(V \times \mathbb{F})$ *is either ordinary or ideal but not both.*■

Finally, we include an analog of Theorem II.17:

THEOREM III.11. *A set* $H \subset \mathbb{FP}^n$ *is an* $(n-1)$*-plane if and only if there exist* $\mathbf{a}_1, \cdots, a_{n+1} \in \mathbb{F}$ *not all zero such that* $\mathbf{x} \in H$ *if and only if there exist homogeneous coordinates* $(x_0, \cdots, x_n)$ *for* $\mathbf{x}$ *such that* $\sum_i a_i x_i = 0$.

Proof. The theorem follows because a subset of $\mathbb{F}^{n+1}$ is an n-dimensional vector subspace if and only if it is the set of all points which solve a nontrivial linear homogeneous equation of the type described in the theorem (compare Theorem A.10).■

Several additional and important properties of the 1–1 mapping

$$\mathrm{J}_V : V \longrightarrow \mathbb{P}(V) = \mathcal{S}_1(V \times \mathbb{F})$$

are developed in the exercises, and the latter are particularly worthy of attention.

EXERCISES

1. Translate the following statements about $\mathbb{F}^3$ into the language of ideal points.

(a) Through a given point there is a unique plane parallel to a given plane.

(b) Two lines which are parallel to a third line are parallel to each other.

(c) If a line is parallel to each of two intersecting planes it is parallel to their line of intersection, and conversely.

(d) If a line L is parallel to a plane P any plane containing L cuts P in a line parallel to L.

(e) Through a given line one plane, and only one, can be passed parallel to a given skew line.

(f) Through a given point one plane, and only one, can be passed that is parallel to each of two skew lines.

(g) All the lines through a point and parallel to a given plane lie on a plane parallel to the first plane.

(h) If a plane contains one of two parallel lines but not the other, it is parallel to the other line.

(i) The intersection of a plane with two parallel planes is a pair of parallel lines.

2. Let V be a vector space over $\mathbb{F}$ of dimension ≥ 3, let $\mathbf{x}_0, \cdots, \mathbf{x}_m$ be distinct points in $\mathcal{S}_1(V)$, and for each i let $\widetilde{\mathbf{x}_i}$ be a set of homogeneous coordinates for $\mathbf{x}_i$. Prove that the set $\{\mathbf{x}_0, \cdots, \mathbf{x}_m\} \subset \mathcal{S}_1(V)$ is *geometrically* independent if and only if the set $\{\widetilde{\mathbf{x}_0}, \cdots, \widetilde{\mathbf{x}_m}\} \subset V$ is *linearly* independent.

3. Determine which of the following vectors in $\mathbb{R}^3$ correspond to homogeneous coordinates of collinear points in $\mathbb{RP}^2$:

(i) (5, 2, 4) (5, -2, 1) (15, 2, 9)

(*ii*) (3, 1, -2) (8, -3, 4) (5, -2, 6)

(*iii*) (1, 5, 1) (1, 1, -1) (3, 4, 1)

(*iv*) (1, 2, 3) (3, 0, 3) (-2. 3, -1)

4. Determine which of the following vectors in $\mathbb{R} * 4$ correspond to homogeneous coordinates of collinear or coplanar points in $\mathbb{RP}^3$:

(*i*) (1, 2, 1, 3) (2, 1, 3, 3) (1, 0, 3, 0) (2, 1, 1, 5)

(*ii*) (1, 1, 2, 1) (0, 1, 1, 2) (-1, 1, 2, 0) (2, 0, 0, -3)

(*iii*) (1, 1, 1, 0) (1, 1, 0, 1) (1, 0, 1, 1) (0, 1, 1, 1)

(*iv*) (1, 2, 1, 0) (1, 0, -1, -1) (0, 2, 2, 1) (1, 2, 0, -1)

5. Let $\mathbb{F}$ be a field with q elements (for example, if p is a prime then the field $\mathbb{Z}_p$ has p elements). Prove that $\mathbb{FP}^n$ has $1 + q + \cdots + q^n$ points. [*Hint:* Let $\pi : \mathbb{F}^{n+1} - \mathbf{0} \to \mathbb{FP}^n$ be the map taking a nonzero vector to the point for which it is a set of homogeneous coordinates. Explain why there are $(q-1)$ possible choices of homogeneous coordinates for every point. Using simple counting considerations, show that the number k of points in $\mathbb{FP}^n$ times the number of choices for homogeneous coordinates is equal to the number of nonzero elements in $\mathbb{F}^{n+1}$, which is $q^{n+1} - 1$. This yields an equation for k; solve this equation.]

6. For each of the pairs of planes in $\mathbb{R}^3$ given below, the intersection is a line L in $\mathbb{R}^3$. Find homogeneous coordinates for the ideal point of L.

(*i*) $3x + 3y + z = 2$ and $3x - 2y = -13$

(*i*) $1x + 2y + 3z = 4$ and $2x + 7y + z = 8$

7. Prove Theorem 11.

> HYPOTHESIS AND NOTATION. For the rest of these exercises, assume that V is a finite-dimensional (left or right) vector space over $\mathbb{F}$. If $\mathbf{x} \in V$ is nonzero, we shall denote the 1-dimensional subspace spanned by $\mathbf{x}$ by $\mathbb{F} \cdot \mathbf{x}$ or more simply by $\mathbb{F}\mathbf{x}$.

8. Using Theorem 9, answer the following questions and prove that your answer is correct:

(*i*) In $\mathcal{S}_1(V)$, what is the intersection of a line with a hyperplane that does not contain it?

(*ii*) In $\mathcal{S}_1(V)$, what is the smallest number of hyperplanes that do not contain a common point?

9. Prove that the projective extension map

$$\mathrm{J}_V : V \longrightarrow \mathbb{P}(V) = \mathcal{S}_1(V \times \mathbb{F})$$

is 1–1 but not onto.

10. Let Q be a k-plane in V (where $k \leq \dim V$). Prove that $J[Q]$ is contained in a unique k-plane $Q' \subset \mathbb{P}(V)$. [*Hint:* If $\{\mathbf{x}_0, \cdots, \mathbf{x}_k\}$ is an affine basis for Q, let W be the affine span of the vectors $(\mathbf{x}_i, 1)$ and consider $\mathbb{P}(W)$.]

11. In the preceding exercise, prove that the ideal points of $\mathbb{P}(W)$ form a $(k-1)$-plane in $\mathbb{P}(V \times \mathbb{F})$. In particular, every line contains a unique ideal point and every plane contains a unique line of ideal points.

12. (*Partial converse to Exercise* 10). Let W be a $(k+1)$-dimensional subspace of $V \times \mathbb{F}$ such that

$$\dim W \ - \ \dim\bigl(W \cap (V \times \{\mathbf{0}\}\bigr) \ = \ 1 \ .$$

Prove that $J^{-1}\bigl[\,\mathbb{P}(W)\,\bigr]$ is a k-plane of V.

13. Let V and W be vector spaces over $\mathbb{F}$, and let $T : V \to W$ be a linear transformation that is one-to-one.

(*i*) Let $\mathbf{v} \neq \mathbf{0}$ in V. Show that the mapping $\mathcal{S}_1(T) : \mathcal{S}_1(V) \to \mathcal{S}_1(W)$ taking $\mathbb{F} \cdot \mathbf{v}$ to $\mathbb{F} \cdot T(\mathbf{v})$ is well-defined (*i.e.*, if $\mathbf{v}$ and $\mathbf{x}$ are nonzero vectors that are nonzero multiples of each other than so are their images under T).

(*ii*) If $S : W \to X$ is also a linear transformation, show that $\mathcal{S}_1(S \circ T) \ = \ \mathcal{S}_1(S) \circ \mathcal{S}_1(T)$. Also show that if T is an identity mapping then so is $\mathcal{S}_1(T)$.

(*iii*) If T is invertible, prove that $\mathcal{S}_1(T)$ is also invertible and that $\mathcal{S}_1(T)^{-1} \ = \ \mathcal{S}_1\left(T^{-1}\right)$.

14. Suppose that V is an n-dimensional vector space and let $\{\,\mathbf{x}_1, \ \cdots \ , \mathbf{x}_n\,\}$ and $\{\,\mathbf{y}_1, \ \cdots \ , \mathbf{y}_n\,\}$ be geometrically independent subsets of $\mathcal{S}_1(V)$. Find an invertible linear transformation T such that $\mathcal{S}_1(T)$ maps $\mathbf{x}_i$ to $\mathbf{y}_i$ for $1 \leq i \leq n$.

15. Let T_0 be an element of the affine group $\mathrm{Aff}(V)$. Prove that there is an invertible linear transformation $T : V \times \mathbb{F} \to V \times \mathbb{F}$ such that

$$\mathcal{S}_1(T) \circ \mathrm{J}_V \ = \ \mathrm{J}_V \circ T_0 \ .$$

[*Hint:* Let $\{\,\mathbf{v}_1, \ \cdots \ , \mathbf{v}_n\,\}$ be a linear basis for V and choose T so that $T(\mathbf{v}_i, 0) = (\,T_0(\mathbf{v}_i), 0\,)$ for all i and $T(\mathbf{0}, 1) = (\mathbf{0}, 1)$. Show that $T(\mathbf{x}, 1) = (\varphi(\mathbf{x}), 1)$ for all $\mathbf{x} \in V$ using the expansion of $\mathbf{x}$ as an affine combination of the basis vectors $\{\,\mathbf{v}_1, \ \cdots \ , \mathbf{v}_n\,\}$ and the zero vector.]

16. Suppose that $T : V \times \mathbb{F} \to V \times \mathbb{F}$ is an invertible linear transformation which maps $V \times \{\mathbf{0}\}$ to itself. Prove that $\mathcal{S}_1(T)$ maps $\mathrm{J}_V[V]$ to itself and the induced self-map of V is in $\mathrm{Aff}(V)$.

17. Let S and T be geometrical subspaces of $\mathcal{S}_1(V)$. Prove that the join $S \star T$ is the set of all points $\mathbf{z}$ such that $\mathbf{z} \in S \cup T$ or $\mathbf{z} \in \mathbf{xy}$, where $\mathbf{x} \in S$ and $\mathbf{y} \in T$. In other words, the join $S \star T$ is the set of all points on lines **joining** points of S and T. [*Hint:* If $S = \mathcal{S}_1(W)$ and $T = \mathcal{S}_1(U)$, then $S \star T \ = \ \mathcal{S}_1(W + U)$.]

> *Note.* A corresponding description of the join in *affine* geometry is given in Appendix B.

Addendum. Synthetic construction of projective space

The preceding three sections of this chapter described the analytic approach to generalizing the synthetic construction of Section III.1. In this addendum to Section III.4 we shall discuss the synthetic approach to construction of a projective space from an affine space. Since some of the arguments are lengthy and we shall not use this material subsequently except in a few peripheral exercises and remarks, many of the details have been omitted; the latter are generally straightforward (but often tedious), and thus they are left to the reader.

We shall need the following results on affine incidence spaces of arbitrary dimension. The proofs are quite similar to their 3-dimensional special cases in Section II.1.

THEOREM III.12. *Suppose that L, M and N are lines in an affine incidence n-space ($n \geq 2$) and $L \neq N$. If $L||M$ and $M||N$, then $L||N$.*■

THEOREM III.13. *Let S be an an affine n-space let H be a hyperplane in S, and let $\mathbf{x}$ be a point of S which does not lie in H. Then there is a unique hyperplane H' in S such that $\mathbf{x} \in H'$ and $H \cap H' = \varnothing$ (in other words, a* **parallel hyperplane** *to H which passes through $\mathbf{x}$).*■

Motivated by the 2-dimensional case, define a *direction* in an affine n-space S to be an equivalence class of parallel lines, and set $S^\wedge$ equal to S together with all the directions in S. Denote the set of all directions by S_∞. We define tso types of geometrical subspaces of $S^\wedge$ as follows:

(A) *Extensions of subspaces of S.* If P is a geometrical subspace of S, set $P^\wedge$ equal to P together with all directions $L^\wedge$ containing a representative L_0 which lies in P. The **dimension** of $P^\wedge$ is defined to be equal to the dimension of P.

(B) *Ideal geometrical subspaces* or *subspaces at infinity.* If P is a geometrical subspace of S, then its set of ideal points $P_\infty = P^\wedge \cap S_\infty$ is a geometrical subspace whose dimension is equal to $d(P) - 1$. [In particular, S_∞ is a hyperplane in S, and it is called the *ideal hyperplane* or the *hyperplane (of points) at infinity.*

We shall take the above as the **definition** of the *synthetic projective extension $S^\wedge$ of an affine n-space S*, and $\mathrm{X}_S : S \to S^\wedge$ will denote the inclusion of S in its projective extension.

The following result expresses the equivalences between the synthetic and analytic approaches to projective extensions of coordinate affine spaces.

THEOREM III.14. *Let V be a finite-dimensional vector space over $\mathbb{F}$ such that $\dim V \geq 2$. Then there is a $1-1$ correspondence*

$$h_V : V^\wedge \longrightarrow \mathcal{S}_1(V \times \mathbb{F})$$

with the following properties:

(i) $\mathrm{J}_V = h_V \circ \mathrm{X}_V$

(ii) A subset $Q \subset V^\wedge$ is a k-plane in $V^\wedge$ if and only if $h_V[Q]$ is a k-plane in $\mathcal{S}_1(V \times \mathbb{F})$.

Proof. We shall break the argument down into a sequence of steps.

STEP 1. *Construction of a* $1-1$ *correspondence between* $V^\wedge$ *and* $\mathcal{S}_1(V \times \mathbb{F})$.

If $\mathbf{v} \in V$ and $\mathrm{X}(\mathbf{v}) = \mathbf{v}'$, define $h_V(\mathbf{v}') = \mathbb{F} \cdot (\mathbf{v},\, 1)$; this construction is well-defined since the mapping X_V is 1–1. Given an ideal point $L^\wedge$ associated to some line L, let $\mathbf{u} \neq \mathbf{0}$ span the unique 1-dimensional vector subspace of V which is parallel or equal to L, and let

$$h_V\left(L^\wedge\right) \;=\; \mathbb{F} \cdot (\mathbf{u},\, 0)\ .$$

This is well-defined because $L||M$ implies that $\mathbf{0u}$ is parallel or equal to M (by Theorem 13), and the right hand side of the formula remains the same if we replace $\mathbf{u}$ by any other nonzero point $\mathbf{u}'$ of $\mathbf{0u}$ (since $\mathbf{u}'$ and $\mathbf{u}$ are nonzero multiples of each other). By definition we have $\mathrm{J}_V = h_V \circ \mathrm{X}_V$.

We shall now show that h_V is 1–1. (a) If $\mathbf{a},\ \mathbf{b} \in V$, are distinct vectors, then $\mathrm{X}_V(\mathbf{a}) \neq \mathrm{X}_V(\mathbf{b})$. If the images of these points under h_V are equal, then we have

$$\mathrm{J}_X(\mathbf{a}) \;=\; h_V \circ \mathrm{X}_V(\mathbf{a}) \;=\; h_V \circ \mathrm{X}_V(\mathbf{b}) \;=\; \mathrm{J}_V(\mathbf{b})$$

which is a contradiction because J_V is known to be 1–1. Therefore the map h_V is 1–1 on the image of X_V. (b) Suppose now that $L^\wedge$ and $M^\wedge$ are ideal points such that

$$h_V\left(L^\wedge\right) \;=\; h_V\left(M^\wedge\right)\ .$$

This means that there is a single line $\mathbf{0u}$ in V which is parallel or equal to each of L and M. which in turn means that $L^\wedge \;=\; M^\wedge$. (c) Suppose now that we have an ordinary point $\mathrm{X}_V(\mathbf{a})$ and an ideal point $L^\wedge$ which have the same image under h_V. If this is true then $\mathbb{F} \cdot (\mathbf{a},\, 1)$ is equal to $\mathbb{F} \cdot (\mathbf{u},\, 0)$, where $\mathbf{u}$ is given as before. Since every nonzero vector in the first subspace has a nonzero last coordinate and every vector in the second subspace has a zero last coordinate, it is clear that the two subspaces cannot be equal, and therefore an ordinary point and an ideal point cannot have the same images under h_V. — This completes the proof that h_V is 1–1.

We shall now show that h_V is onto. Let $\mathbf{0} \neq (\mathbf{v},\, c) \in V \times \mathbb{F}$. If $c \neq 0$, then we have

$$\mathbb{F} \cdot (\mathbf{v},\, c) \;=\; \mathbb{F} \cdot (c^{-1}\, \mathbf{v},\, 1) \;=\; \mathrm{J}_V(c^{-1}\mathbf{v}) \;=\; h_V \circ XX_V(c^{-1}\mathbf{v})\ .$$

On the other hand, if $c = 0$ then $\mathbf{v} \neq \mathbf{0}$ and

$$\mathbb{F} \cdot (\mathbf{v}, 0) \;=\; h_V\left((\mathbf{0v})^\wedge\right)\ .$$

STEP 2. *Under the above correspondence, a subset* $Q \subset V^\wedge$ *is a hyperplane if* $h_V[Q]$ *is a hyperplane in* $\mathcal{S}_1(V \times \mathbb{F})$. — We shall only sketch the argument, leaving verification of the details to the reader.

This uses a result from the Exercises for Section 4 of Appendix A: *If* X *is* m*-dimensional vector space over* $\mathbb{F}$*, then* $Y \subset X$ *is an* $(m-1)$*-dimensional vector subspaces if and only if there is a nonzero linear transformation (or* functional*)* $g : X \to \mathbb{F}$ *such that* $\mathbf{x} \in Y$ *if and only if* $g(\mathbf{x}) = 0$. — The linear functional g is not unique, for if Y is the zero set for g and c is a nonzero constant, then Y is also the zero set for $c \cdot g$.

By the previously mentioned exercises and the discussion following Theorem A.12, every linear functional on $V \times \mathbb{F}$ has the form $g(\mathbf{v},\, t) \;=\; g_0(\mathbf{v}) + a \cdot t$, where g_0 is a linear functional on V and $a \in \mathbb{F}$. — Extensions of ordinary hyperplanes in V are defined by expressions of this form for which $g_0 \neq 0$, and the ideal hyperplane in V is defined by linear functionals of this form in which $g_0 = 0$ and $b \neq 0$.

Suppose we are given a hyperplane $H^\wedge$ in $V^\wedge$ which is the extension of a hyperplane H in V. Then H is defined by an equation of the form $f(\mathbf{v}) = b$ for some nonzero linear functional f and scalar b, and $h_V\,[\,H^\wedge\,]$ is equal to $\mathcal{S}_1(W)$, where W is the zero set of the linear functional $g(\mathbf{v},\,t) \;=\; f(\mathbf{v}) - b\cdot t$. Since we also know that h_V maps the ideal hyperplane in $V^\wedge$ to the zero set of the functional $g(\mathbf{v},\,t) \;=\; t$, it follows that the image of a hyperplane in $V^\wedge$ is a hyperplane in $\mathcal{S}_1(V \times \mathbb{F})$.$\square$

Conversely, suppose that $\mathcal{S}_1(W)$ is a hyperplane in $\mathcal{S}_1(V\times\mathbb{F})$, and suppose that the n-dimensional vector subspace W is defined by the nonzero linear functional. Write $g(\mathbf{v},\,t) \;=\; g_0(\mathbf{v}) + a\cdot t$ as above, where either $g_0 \neq 0$ or $a \neq 0$. In the first of these cases $\mathcal{S}_1(W)$ is the image of the extended ordinary hyperplane defined by the equation $g_0(\mathbf{v}) = -a$, and in the second case $\mathcal{S}_1(W)$ is the image of the ideal hyperplane.

STEP 3. *Let k be an integer satisfying $1 \leq k \leq n-2$. In both $V^\wedge$ and $\mathbb{FP}^n$, a k-dimensional geometrical subspace Q is the intersection of an $(k+1)$-dimensional subspace Q' with a hyperplane H not containing it.*

The proofs of these statements are variants of Exercise II.5.3 which shows that every geometrical subspace is an intersection of hyperplanes. There is an analogous result for vector subspaces of an m-dimensional vector spaces: Every k-dimensional subspace is an intersection of $(m-k)$ distinct hyperplanes. Once again, the argument breaks down into various cases.

STEP 4. *By downward induction on k such that $1 \leq k \leq (n-1)$, one shows that Q is a k-plane in both $V^\wedge$ if and only if $h_V[Q]$ is a k-plane in $\mathbb{FP}^n$.*

Once again, we shall only sketch the argument: The case $k = (n-1)$ is covered by Step 2, and the recursive step — showing that the validity of the result for $(k+1)$ implies its validity for k — follows from the description of k-planes in Step 3.$\blacksquare$

Abstract projective extensions

Definition. Let S be a geometrical incidence space of dimension $n \geq 3$. An *abstract projective extension* of S is a 1–1 map $\varphi : S \to \mathbb{FP}^n$, where $\mathbb{F}$ is some skew-field, such that if Q is a k-plane in S then there is a unique k-plane $Q' \subset \mathbb{FP}^n$ such that $Q = \varphi^{-1}[Q']$.

The results of this chapter prove the existence of projective extensions for the coordinate affine n-spaces (in fact, for *all* affine n-spaces by Theorem II.38). On the other hand, it is not difficult to see that projective extensions exist for many other geometrical incidence spaces (*e.g.*, this holds for the examples in Exercises II.5.7 and II.5.8). In fact, if one has a 3-dimensional (regular) geometrical incidence space which also has a notion of betweenness satisfying Hilbert's Axioms of Order, then the existence of an abstract projective extension (with a suitable analog of betweenness) is given by results due to A. N. Whitehead.[9] Extremely general, and purely incidence-theoretic, conditions for the existence of projective extensions have been established

[9]See Chapter III of the book, *The Axioms of Descriptive Geometry*, in the bibliography. — ALFRED NORTH WHITEHEAD (1861–1947) was an extremely well-known philosopher who worked extensively on the logical foundations of mathematics during the period from the late 1880s until about 1913, at which time he shifted his attention to other areas of philosophy. Whitehead is particularly known for his study of the foundations of mathematics with Bertrand Russell (1872–1970), which is largely contained in a massive and ambitious three volume work called *Principia Mathematica*.

by S. Gorn.[10] A discussion of the classical Non-Euclidean geometries using their projective extensions appears in Chapters 12 and 13 of the book by Fishback listed in the bibliography.

THE 2-DIMENSIONAL CASE. Clearly there is no problem in formulating a definition of abstract projective extension for incidence planes, and the constructions of this chapter shows that every *affine* coordinate incidence plane admits a projective extension. In fact, if one defines an abstract notion of projective plane as in Section IV.1 below, then the constructions of Section III.1 show that for every abstract affine plane P there admits a map $\varphi : P \to P^*$ such that P^* is an abstract projective plane and (as before) if L is a line in P then there is a unique line $L' \subset \mathbb{FP}^n$ such that $L = \varphi^{-1}[L']$. However, two major differences are that (i) in contrast to the situation in higher dimensions, we cannot always take P^* to be a coordinate projective plane $\mathbb{FP}^2$, (ii) the proofs of the results of Whitehead and Gorn on projective extensions for non-affine incidence structures do **not** extend to the 2-dimensional case. We shall discuss these points further in the next chapter.

EXERCISES

1. Assuming that S is an affine n-space, prove that $S^\wedge$ is a regular geometrical incidence space, and in fact the dimensions of subspaces satisfy the strong form of the regularity condition

$$d(T_1 \star T_2) \ = \ d(T_1) \ + \ d(T_2) \ - \ d(T_1 \cap T_2)$$

for all geometrical subspaces T_1 and T_2.

2. Suppose that $f : S \to S'$ is a geometric symmetry of an affine n-space. Prove that f extends to a unique geometric symmetry $f^\wedge$ of $S^\wedge$. [*Hint:* If $L||L'$, then $f[L]||f[L']$; verify this and use it to define $f^\wedge$.]

3. In the notation of the preceding exercise, prove that $(\mathrm{id}_S)^\wedge$ is the identity on $S^\wedge$ and if $g : S \to S$ is another symmetry then $(g \circ f)^\wedge \ = \ g^\wedge \circ f^\wedge$. Finally, show that

$$(f^{-1})^\wedge \ = \ \left(f^\wedge\right)^{-1} \ .$$

[*Hint:* In each desired identity, explain why both sides of the equation extend the same geometric symmetry of S.]

4. Prove Theorems 13 and 14, and fill in the details for the proof of Theorem 15.

---

[10] A reference to the original 1940 research article appears in the bibliography. — SAUL GORN (1913–1992) began his professional career as a mathematician, but his interests moved to computer science with the emergence of that subject during the 1940s, and he played a significant role in the establishment of computer science as an independent branch of the mathematical sciences. As a researcher, he is best known for his theory of mechanical languages based upon work of twentieth century philosophers like L. Wittgenstein (1889–1951) on human linguistics.

CHAPTER IV

SYNTHETIC PROJECTIVE GEOMETRY

The purpose of this chapter is to begin the study of projective spaces, mainly from the synthetic point of view but with considerable attention to coordinate projective geometry.

1. Axioms for projective geometry

The basic incidence properties of coordinate projective spaces are expressible as follows:

Definition. A geometrical incidence space (S, Π, d) is **projective** if the following hold:

(P-1) : Every line contains at least three points.

(P-2) : If P and Q are geometrical subspaces of S then

$$d(P \star Q) \;=\; d(P) \;+\; d(Q) \;-\; d(P \cap Q)\ .$$

In particular, **(P-2)** is a strong version of the regularity condition **(G-4)** introduced in Section II.5. The above properties were established for $\mathbb{FP}^n$ $(n \geq 2)$ in Theorems III.10 and III.9 respectively. It is useful to assume condition **(P-1)** for several reasons; for example, lines in Euclidean geometry have *infinitely many* points, and **(P-1)** implies a high degree of regularity on the incidence structure that is not present in general (compare Exercise 2 below and Theorem IV.11). — In this connection, note that Example 2 in Section II.5 satisfies **(P-2)** and *every* line in this example contains *exactly* two points.

Elementary properties of projective spaces

The following is a simple consequence of the definitions.

THEOREM IV.1. *If S is a geometrical subspace of a geometrical incidence space S', then S is a geometrical incidence space with respect to the subspace incidence structure of Exercise* II.5.3.■

If P is a projective incidence space and $d(P) = n \geq 1$, then P is called a *projective n-space*; if $n = 2$ or 1, then one also says that P is a *projective plane* or *projective line*, respectively.

THEOREM IV.2. *If P is a projective plane and L and M are distinct lines in P, then $L \cap M$ consists of a single point.*■

THEOREM IV.3. *If S is a projective 3-space and P and Q are distinct planes in S, then $P \cap Q$ is a line.*■

These follow from **(P-2)** exactly as Examples 1 and 2 in Section III.4 follow from Theorem III.9.

We conclude this section with another simple but important result:

THEOREM IV.4. *In the definition of a projective space, property* **(P-1)** *is equivalent to the following (provided the space is not a line):*

> **(P-1′) :** *Every plane contains a subset of four points, no three of which are collinear.*

Proof. Suppose that **(P-1)** holds. Let P be a plane, and let X, Y and Z be noncollinear points in P. Then the lines $L = XY$, $M = XZ$, and $N = YZ$ are distinct and contained in P. Let W be a third point of L, and let V be a third point of M.

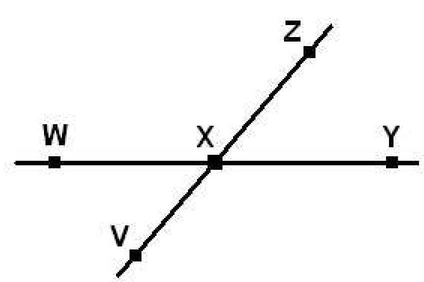


Figure IV.1

Since L and M are distinct and meet at X, it follows that the points V, W, Y, Z must be distinct (if any two are equal then we would have $L = M$; note that there are six cases to check, with one for each pair of letters taken from W, X, Y, Z). Similarly, if any three of these four points were collinear then we would have $L = M$, and therefore no three of the points can be collinear (there are four separate cases that must be checked; these are left to the reader).□

Conversely, suppose that **(P-1′)** holds. Let L be a line, and let P be a plane containing L. By our assumptions, there are four points A, B, C, $D \in P$ such that no three are collinear.

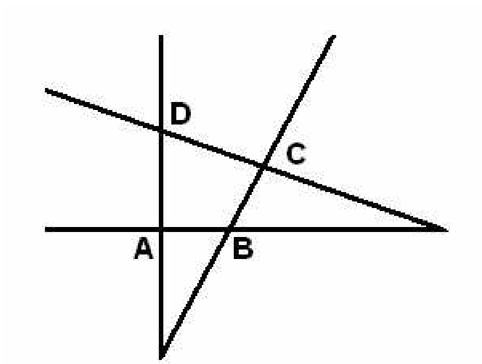


Figure IV.2

Let $M_1 = AB$, $M_2 = BC$, $M_3 = CD$, and $M_4 = AD$. Then the lines M_1 are distinct and coplanar, and no three of them are concurrent (for example, $M_1 \cap M_2 \neq M_3 \cap M_4$, and similarly for the others). It is immediate that M_1 contains at least three distinct points; namely, the points A and B plus the point where M_1 meets M_3 (these three points are distinct because no three of the lines M_i are concurrent). Similarly, each of the lines M_2, M_3 and M_4 must contain at least three points.

If L is one of the four lines described above, then we are done. Suppose now that L is **not** equal to any of these lines, and let P_i be the point where L meets M_i. If at least three of the points $P_1,\ P_2,\ P_3,\ P_4$ are distinct, then we have our three distinct points on L. Since no three of the lines M_i are concurrent, it follows that no three of the points P_i can be equal, and therefore if there are not three distinct points among the P_i then there must be two distinct points, with each P_i equal to a unique P_j for $j \neq i$. Renaming the M_i if necessary by a suitable reordering of $\{1, 2, 3, 4\}$, we may assume that the equal pairs are given by $P_1 = P_3$ and $P_2 = P_4$. The drawing below illustrates how such a situation can actually arise.

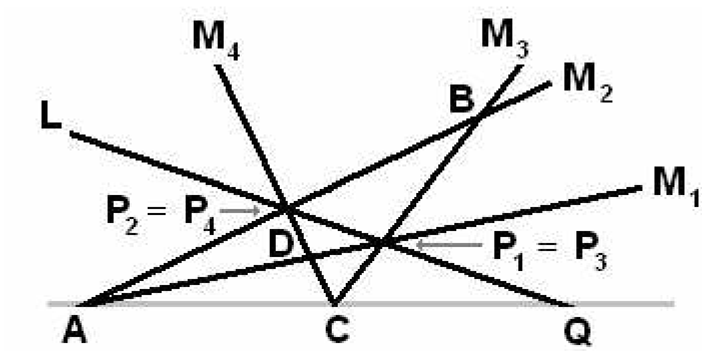


Figure IV.3

We know that $P_1 = P_3$ and $P_2 = P_4$ are two distinct points of L, and Figure IV.3 suggests that the point Q where AC meets L is a third point of L. To **prove** this, we claim it will suffice to verify the following statements motivated by Figure IV.3:

(i) The point A does not lie on L.

(ii) The line AC is distinct from M_1 and M_2.

Given these properties, it follows immediately that the three lines AC, M_1 and M_2 — which all pass through the point A which does not lie on L — must meet L in three distinct points (see Exercise 4 below).

Assertion (i) follows because $A \in L$ implies

$$A \ \in \ M_2 \cap L \ = \ M_4 \ \cap \ M_2 \ \cap \ L$$

and since $A \in M_1 \cap M_2$ this means that M_1, M_2, and M_4 are concurrent at A. However, we know this is false, so we must have $A \in L$. To prove assertion (ii), note that if $AC = M_1 = AB$, then $A,\ B,\ C$ are collinear, and the same conclusion will hold if $AC = M_2 = BC$. Since the points $A,\ B,\ C$ are noncollinear by construction, it follows that (ii) must also hold, and as noted above this completes the proof that L has at least three points.■

EXERCISES

1. Let (S, Π, d) be an n-dimensional projective incidence space ($n \geq 2$), let P be a plane in S, and let $X \in P$. Prove that there are at least three distinct lines in P which contain X.

2. Let $n \geq 3$ be an integer, let P be the set $\{0,\ 1,\ \cdots\ , n\}$, and take the family of subsets $\mathcal{L}$ whose elements are $\{1,\ \cdots\ , n\}$ and all subsets of the form $\{0,\ k\}$, where $k > 0$. Show that $(P, \mathcal{L})$ is a regular incidence plane which satisfies **(P-2)** but not **(P-1)**. [*Hint:* In this case **(P-2)** is equivalent to the conclusion of Theorem IV.2.]

3. This is a generalization of the previous exercise. Let S be a geometrical incidence space of dimension $n \geq 2$, and let ∞_S be an object not belonging to S (the axioms for set theory give us explicit choices, but the method of construction is unimportant). Define the **cone** on S to be $S^\bullet = S \cup \{\infty_S\}$, and define a subset Q of $S^\bullet$ to be a k-planes of $S^\bullet$ if and only if **either** Q is a k-plane of S **or** $Q = Q_0 \cup \{\infty_S\}$, where Q_0 is a $(k-1)$-plane in S (as usual, a 0-plane is a set consisting of exactly one element). Prove that $S^\bullet$ with these definitions of k-planes is a geometrical incidence $(n+1)$-space, and that $S^\bullet$ satisfies **(P-2)** if and only if S does. Explain why $S^\bullet$ does not satisfy **(P-1)** and hence is not projective.

4. Let $(P, \mathcal{L})$ be an incidence plane, let L be a line in P, let X be a point in P which does not lie on L, and assume that $M_1,\ \cdots\ M_k$ are lines which pass through X and meet L in points $Y_1,\ \cdots\ Y_k$ respectively. Prove that the points $Y_1,\ \cdots\ Y_k$ are distinct if and only if the lines $M_1,\ \cdots\ M_k$ are distinct.

5. Let (S, Π, d) be a regular incidence space of dimension ≥ 3, and assume that every plane in S is projective (so it follows that **(P-1)** holds). Prove that S is projective. [*Hint:* Since S is regular, condition **(P-2)** can only fail to be true for geometrical subspaces Q and R such that $Q \cap R = \varnothing$. If $d(R) = 0$, so that R consists of a single point, then condition **(P-2)** holds by Theorem II.30. Assume by induction that **(P-2)** holds whenever $d(R_0) \leq k-1$, and suppose that $d(R) = k$. Let $R_0 \subset R$ be $(k-1)$-dimensional, and choose $\mathbf{y} \in R$ such that $\mathbf{y} \notin R_0$. Show that $Q \star R_0 \subset Q \star R$, and using this prove that $d(Q \star R)$ is equal to $d(Q) + k$ or $d(Q) + k + 1$. The latter is the conclusion we want, so assume it is false. Given $\mathbf{x} \in Q$, let $\mathbf{x}\,R$ denote the join of $\{\mathbf{x}\}$ and R, and define $\mathbf{y}\,Q$ similarly. Show that $\mathbf{x}\,R \cap Q$ is a line that we shall call L, and also show that $R \cap \mathbf{y}\,Q$ is a line that we shall call M. Since $L \subset Q$ and $M \subset R$, it follows that $L \cap M = \varnothing$. Finally, show that $\mathbf{x}\,R\ \cap\ \mathbf{y}\,Q$ is a plane, and this plane contains both L and M. Since we are assuming all planes in S are projective, it follows that $L \cap M$ is nonempty, contradicting our previous conclusion about this intersection. Why does this imply that $d(Q \star R)\ =\ d(Q) + k + 1$?]

2. Desargues' Theorem

In this section we shall prove a synthetic version of a fundamental result of plane geometry due to G. Desargues (1591–1661). The formulation and proof of Desargues' Theorem show that projective geometry provides an effective framework for proving nontrivial geometrical theorems.

THEOREM IV.5. (Desargues' Theorem) *Let P be a projective incidence space of dimension at least three, and let $\{A, B, C\}$ and $\{A', B', C'\}$ be triples of noncollinear points such that the lines AA', BB' and CC' are concurrent at some point X which does not belong to either of $\{A, B, C\}$ and $\{A', B', C'\}$. Then the points*

$$\begin{array}{lcccc} D & \in & BC & \cap & B'C' \\ E & \in & AC & \cap & A'C' \\ F & \in & AB & \cap & A'B' \end{array}$$

are collinear.

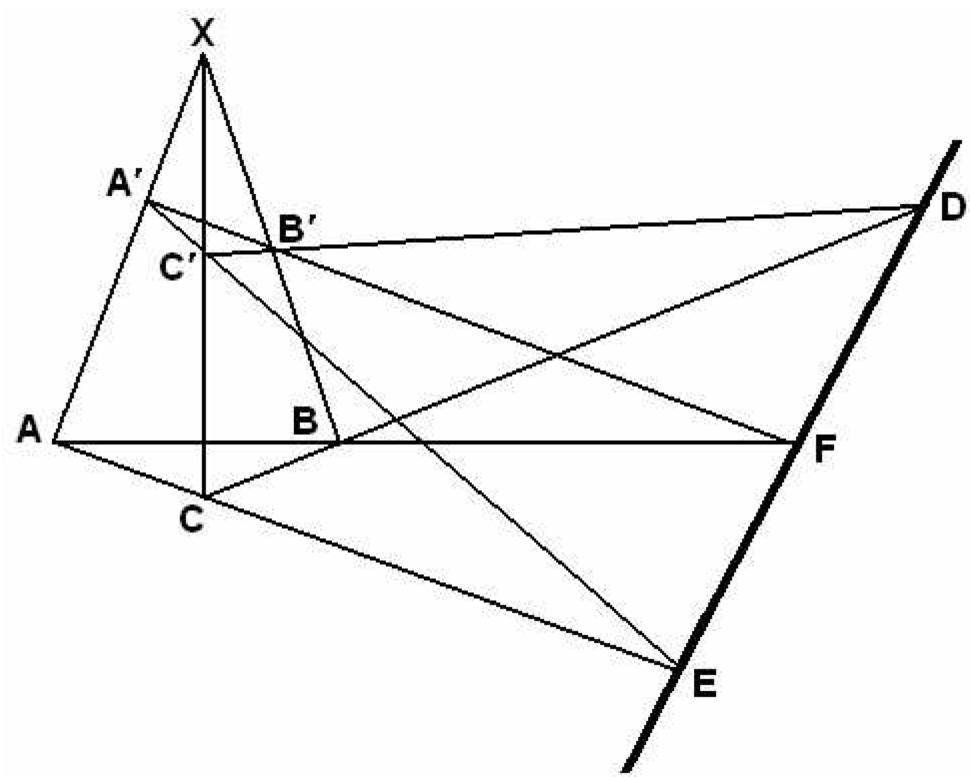


Figure IV.4

Proof. The proof splits into two cases, depending upon whether or not the sets $\{A, B, C\}$ and $\{A', B', C'\}$ are coplanar. One feature of the proof that may seem counter-intuitive is that the noncoplanar case is the easier one. In fact, we shall derive the coplanar case using the validity of the result in the noncoplanar case.

CASE 1. *Suppose that the planes determined by* $\{A, B, C\}$ *and* $\{A', B', C'\}$ *are distinct.* Then the point X which lies on AA', BB' and CC' cannot lie in either plane. On the other hand, it follows that the points $\{A', B', C'\}$ lie in the 3-space S determined by $\{X, A, B, C\}$, and therefore we also know that the planes ABC and $A'B'C'$ are contained in S. These two planes are distinct (otherwise the two triples of noncollinear points would be the same), and hence their intersection is a line. By definition, all three of the points D, E, F all lie in the intersection of the two planes, and therefore they all lie on the two planes' line of intersection.

CASE 2. *Suppose that the planes determined by* $\{A, B, C\}$ *and* $\{A', B', C'\}$ *are identical.* The idea is to realize the given configuration as the photographic projection of a similar noncoplanar configuration on the common plane. Since photographic projections onto planes preserve collinearity, this such a realization will imply that the original three points D, E, F are all collinear.

Under the hypothesis of Case 2, all the points under consideration lie on a single plane we shall call P. Let Y be a point not on P, and let $Z \in AY$ be another. Consider the line $A'Z$; since A' and Z both lie on the plane AXY, the whole line $A'Z$ lies in AXY. Thus $A'Z$ and XY meet in a point we shall call Q.

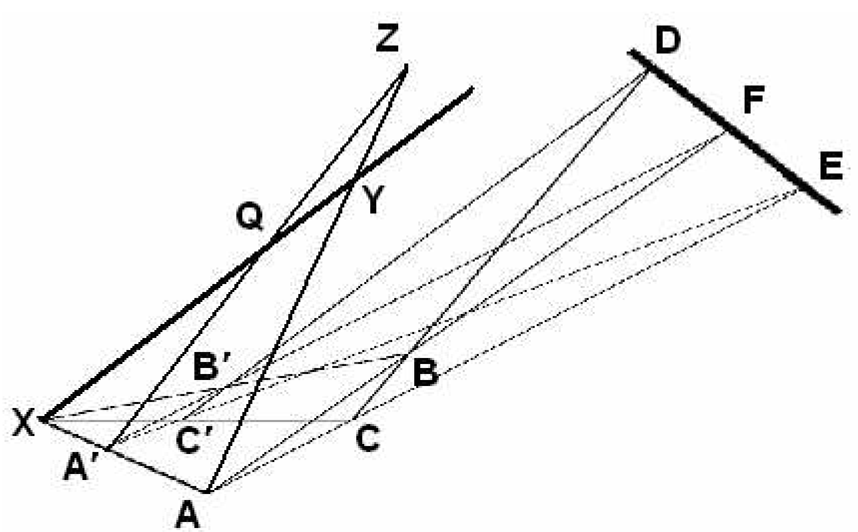


Figure IV.5

Consider the following three noncoplanar triangle pairs:

(i) $C'QB'$ and CYB.

(ii) $C'QA'$ and CYA.

(iii) $B'QA'$ and BYA.

Since AA', BB' and CC' all meet at X, the nonplanar case of the theorem applies in all thre cases. Let $G \in BY \cap B'Q$ and $H \in CY \cap C'Q$, and note that $Z \in AR \cap A'Q$. Then the truth of the theorem in the noncoplanar case implies that each of the triples

$$\{D, H, G\}, \qquad \{F, Y, G\}, \qquad \{E, H, Z\}$$

is collinear.

Let P' be the plane DFZ. Then $G \in P'$ by the collinearity of the second triple, and hence $H \in P'$ by the collinearity of the first triple. Since $Z, H \in P'$, we have $E \in P'$ by the collinearity of the third triple. All of this implies that

$$E \in P' \cap P = DF$$

which shows that the set $\{D, E, F\}$ is collinear.■

Definition. A projective plane P is said to be *Desarguian* if Theorem 5 is always valid in P.

By Theorem 5, every projective plane that is isomorphic to a plane in a projective space of higher dimension is Desarguian. In particular, if $\mathbb{F}$ is a skew-field, then $\mathbb{FP}^2$ is Desarguian because $\mathbb{FP}^2$ is isomorphic to the plane in $\mathbb{FP}^3$ consisting of all points having homogeneous coordinates in $\mathbb{F}^4$ of the form $(x_1, x_2, x_3, 0)$. In Section 4 we shall note that, conversely, every Desarguian plane is isomorphic to a plane in a projective 3-space.

An example of a non-Desarguian projective plane (the *Moulton plane*)[1] can be given by taking the real projective plane $\mathbb{RP}^2$ as the underlying set of points, and modifying the definition of lines as follows: The new lines will include the line at infinity, all lines which have slope ≤ 0 or are parallel to the y-axis, and the broken lines defined by the equations

$$\begin{array}{lll} y = m(x-a), & x \leq a & (i.e.,\ y \leq 0) \\ y = \frac{1}{2}m(x-a), & x \geq a & (i.e.,\ y \geq 0) \end{array}$$

where $m > 0$. As the points at infinite of the latter lines we take those belonging to $y = m(x-a)$. A straightforward argument shows that the axioms for a projective plane are satisfied (see Exercise 4 below). However, as Figure IV.6 suggests, Desargues' Theorem is false in this plane.

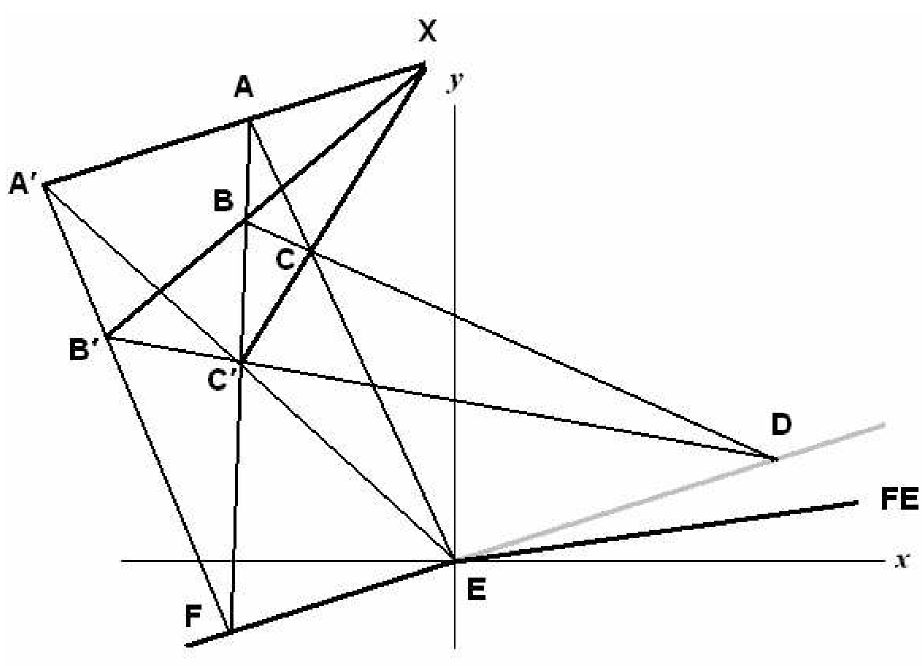


Figure IV.6

[1]Forest Ray Moulton (1872–1952) was an American scientist who worked mostly in astronomy but is also recognized for his contributions to mathematics.

Informally speaking, Desargues' Theorem fails to hold for some projective planes because there is not enough room in two-dimensional spaces to apply standard techniques.[2] Perhaps surprisingly, there are several other significant examples of geometrical problems in which higher dimensional cases (say $n \geq N$, where N depends upon the problem) are simpler to handle than lower dimensional ones (see the paper of Gorn for another example).

Finally, we note that Theorems II.27–29 are basically special cases of Desargues' Theorem.

EXERCISES

1. Explain why the results mentioned above are essentially special cases of Desargues' Theorem.

2. Is it possible to "plant ten trees in ten rows of three?" Explain your answer using Desargues' Theorem.[3]

3. Explain why each of the following pairs of triangles in Euclidean 3-space satisfies the hypotheses of Desargues' Theorem.

(i) Two coplanar triangles such that the lines joining the corresponding vertices are parallel.

(ii) A triangle and the triangle formed by joining the midpoints of its sides.

(iii) Two congruent triangles in distinct planes whose corresponding sides are parallel.

4. Prove that the Moulton plane (defined in the notes) is a projective plane.

[2]For an example of a *finite* Non-Desarguian plane. see pages 158–159 of Hartshorne's book.

[3]For more information on the relevance of projective geometry to counting and arrangement problems, see Section IV.3.

3. Duality

By the principle of duality ... geometry is at one stroke [nearly] doubled in extent with no expenditure of extra labor, — **Eric Temple Bell (1883–1960)**, *Men of Mathematics*[4]

Consider the following fundamental properties of projective planes:

(1) Given two distinct points, there is a unique line containing both of them.

(1*) Given two distinct lines, there is a unique point contained in both of them.

(2) Every line contains at least three distinct points.

(2*) Every point is contained in at least three distinct lines.

The important point to notice is that Statement (n^*) is obtained from Statement (n) by interchanging the following words and phrases:

(i) *point* $\longleftrightarrow$ *line*

(ii) *is contained in* $\longleftrightarrow$ *contains*

Furthermore, Statement (n) is obtained from Statement (n^*) by exactly the same process. Since the four properties (1) – (1*) and (2) – (2*) completely characterize projective planes (see Exercise 1 below), one would expect that points and lines in projective planes behave somewhat symmetrically with respect to each other.

This can be made mathematically precise in the following manner: Given a projective plane $(P, \mathcal{L})$, we define a **dual plane** $(P^*, \mathcal{L}^*)$ such that P^* is the set $\mathcal{L}$ of lines in P and $\mathcal{L}^*$ is in 1–1 correspondence with P. Specifically, for each $x \in P$ we define the *pencil of lines with vertex* x to be the set

$$\mathfrak{p}(x) \;=\; \{\; L \;\in\; \mathcal{L} \;=\; P^* \mid x \;\in\; L \;\} \;.$$

In other words, $L \in \mathfrak{p}(x)$ if and only if $X \in L$. Let P^{**} denote the set of all pencils associated to the projective plane whose points are given by P and whose lines are given by P^*.

THEOREM IV.6. *If* (P, P^*) *satisfies properties* $(1) - (1^*)$ *and* $(2) - (2^*)$*, then* (P^*, P^{**}) *also does.*

Proof. There are four things to check:

(a) Given two lines, there is a unique pencil containing both of them.

(a^*) Given two pencils, there is a unique line contained in both of them.

(b) Every pencil contains at least three lines.

(b^*) Every line is contained in at least three pencils.

---

[4] See the bibliography for more information and comments on this book.

However, it is clear that (a) and (a^*) are rephrasings of (1^*) and (1) respectively, and likewise (b) and (b^*) are rephrasings of (2^*) and (2) respectively. Thus $(1) - (1^*)$ and $(2) - (2^*)$ for (P^*, P^{**}) are logically equivalent to $(1) - (1^*)$ and $(2) - (2^*)$ for (P, P^*).■

As indicated above, we call (P^*, P^{**}) the *dual projective plane* to (P, P^*).

By Theorem 6 we can similarly define P^{***} to be the set of all pencils in P^{**}, and it follows that (P^{**}, P^{***}) is also a projective plane. However, repetition of the pencil construction does not give us anything new because of the following result:

THEOREM IV.7. *Let (P, P^*) be a projective plane, and let $\mathfrak{p}_0 : P \to P^{**}$ be the map sending a point x to the pencil $\mathfrak{p}(x)$ of lines through x. Then $\mathfrak{p}_0$ defines an isomorphism of incidence planes from (P, P^*) to the double dual projective plane (P^{**}, P^{***}).*

Proof. By construction the map $\mathfrak{p}_0$ is onto. It is also 1–1 because $\mathfrak{p}(x) = \mathfrak{p}(y)$ implies every line passing through x also passes through y. This is impossible unless $x = y$.

This it remains to show that L is a line in P if and only if $\mathfrak{p}(L)$ is a line in P^{**}. But lines in P^{**} have the form

$$\mathfrak{p}(L) = \{ w \in P^{**} \mid L \in \mathfrak{p}(x) \} .$$

where L is a line in P. Since $L \in \mathfrak{p}(x)$ if and only if $x \in L$, it follows that $\mathfrak{p}(x) \in \mathfrak{p}(L)$ if and only if $x \in L$. Hence $\mathfrak{p}(L)$ is the image of L under the map $\mathfrak{p}_0$, and thus the latter is an isomorphism of (projective) geometrical incidence planes.■

The preceding theorems yield the following important phenomenon[5] which was described in the first paragraph of this section; it was discovered independently by J.-V. Poncelet (1788–1867) and J. Gergonne (1771–1859).

METATHEOREM IV.8. (Principle of Duality) *A theorem about projective planes remains true if one interchanges the words* point *and* line *and also the phrases* contains *and* is contained in.■

The justification for the Duality Principle is simple. The statement obtained by making the indicated changes is equivalent to *a statement about duals of projective planes* which corresponds to the *original statement for projective planes*. Since duals of projective planes are also projective planes, the modified statement must also hold.

Definition. Let $\mathbf{A}$ be a statement about projective planes. The *dual statement* is the one obtained by making the changes indicated in Metatheorem 8, and it is denoted by $\mathfrak{D}(\mathbf{A})$ or $\mathbf{A}^*$.

EXAMPLE 1. The phrase *three points are collinear* (contained in a common single line) dualizes to *three lines are concurrent* (containing a common single point).

EXAMPLE 2. The property **(P-1′)**, which assumes the existence of four points, not three of which are collinear, dualizes to the statement, *There exist four lines, no three of which are concurrent*. This statement was shown to follow from **(P-1′)** in the course of proving Theorem 4.

[5]This is called a **Metatheorem** because it is really a statement *about* mathematics rather than a theorem *within* mathematics itself. In other words it is a *theorem about theorems*.

By the metatheorem, a statement $\mathbf{A}^*$ is true for all projective planes if $\mathbf{A}^*$ is. On the other hand, Theorem 7 implies that $\mathbf{A}^{**}$ is logically equivalent to $\mathbf{A}$. This is helpful in restating the Principle of Duality in a somewhat more useful form:

Modified Principle of Duality. *Suppose that $\mathbf{A}_1, \cdots, \mathbf{A}_n$ are statements about projective planes. Then $\mathbf{A}_n$ is true in all projective planes satisfying $\mathbf{A}_1, \cdots, \mathbf{A}_{n-1}$ if and only if $\mathbf{A}_n^*$ is true in all projective planes satisfying $\mathbf{A}_1^*, \cdots, \mathbf{A}_{n-1}^*$.*■

In practice, one often knows that the statements $\mathbf{A}_1, \cdots, \mathbf{A}_{n-1}$ imply their own duals. Under these circumstances, the Modified Principle of Duality shows that $\mathbf{A}_n^*$ is true if $\mathbf{A}_n$ is true, and conversely. The next two theorems give significant examples of such statements; in both cases we have $(n-1) = 1$.

THEOREM IV.9. *If a projective plane contains only finitely many points, then it also contains only finitely many lines.*

Proof. Suppose the plane contains n elements. Then there are 2^n subsets of the plane. But lines are subsets of the plane, and hence there are at most 2^n lines.■

In Theorems 11-15 we shall prove very strong duality results for projective planes with only finitely many points. Recall that examples of such systems are given by the coordinate projective planes $\mathbb{Z}_p\mathbb{P}^2$, where p is a prime.

We shall now give a considerably less trivial example involving duality.

THEOREM IV.10. *If a projective plane (P, P^*) is Desarguian, then the dual of Desargues' Theorem is also true in (P, P^*).*

Proof. The first step in the proof is to describe the dual result in terms of (P, P^*).

The dualization of two abstract triples of noncollinear points is two distinct triples of nonconcurrent lines, which we denote by $\{\alpha, \beta, \gamma\}$ and $\{\alpha', \beta', \gamma'\}$. Next, the concurrency hypothesis for AA', BB' and CC' dualizes to a hypothesis that $E \in \alpha \cap \alpha'$, $F \in \beta \cap \beta'$, and $D \in \gamma \cap \gamma'$ are collinear. Finally, the collinearity conclusion in Desargues' Theorem dualizes to a statement that three lines are concurrent. Specifically, if we set

$$\begin{array}{lll} A \in \beta \cap \gamma & \qquad & A' \in \beta' \cap \gamma' \\ B \in \alpha \cap \gamma & & B' \in \alpha' \cap \gamma' \\ C \in \alpha \cap \beta & & C' \in \alpha' \cap \beta' \end{array}$$

then we wish to show that AA', BB' and CC' are concurrent. As the drawing below illustrates, the data for the dual theorem are similar to the data for the original theorem, the key difference being that Desargues' Theorem assumes concurrency of the lines AA', BB' and CC', using this to prove the collinearity of D, E and F, while the dual theorem assumes collinearity of the three points and aims to prove concurrency of the three lines.

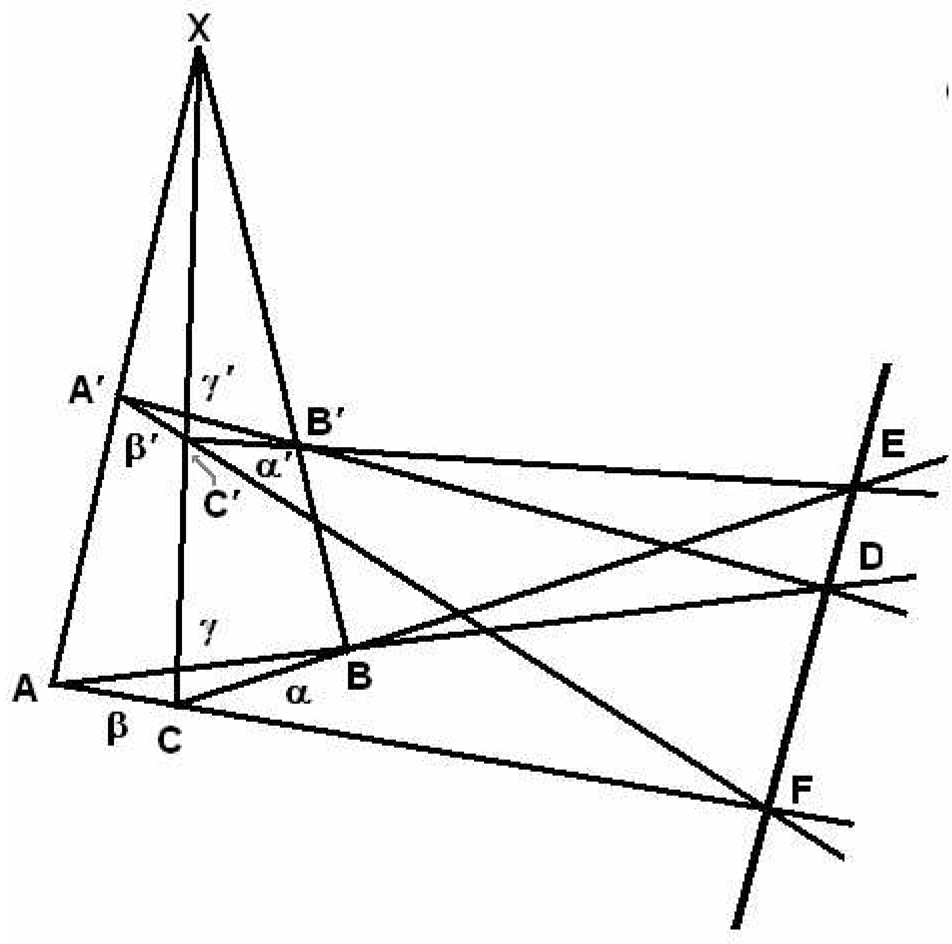


Figure IV.7

By construction, the lines $B'C'$, BC and DF all meet at E. Therefore Desargues' Theorem applies to the triples $\{C', C, F\}$ and $\{B', B, D\}$, and hence we may conclude that the three points

$$\begin{array}{rcl} X & \in & BB' \cap CC' \\ G & \in & BD \cap CF \\ H & \in & B'D \cap C'F \end{array}$$

are collinear. Using this, we obtain the following additional conclusions:

(1) $B,\ D \in \gamma$, and therefore $\gamma = BD$.

(2) $B',\ D \in \gamma'$, and therefore $\gamma' = B'D$.

(3) $C,\ F \in \beta$, and therefore $\beta = CF$.

(4) $C',\ F \in \beta'$, and therefore $\beta' = C'F$.

Therefore $G \in \gamma \cap \beta$ and $H \in \gamma' \cap \beta'$. Since the common points of these lines are A and A' by definition, we see that X, A and A' are collinear. In other words, we have $X \in AA' \cap BB' \cap CC'$, and hence the three lines are concurrent, which is what we wanted to prove.■

Duality and finite projective planes

We shall illustrate the usefulness of duality by proving a few simple but far-reaching results on projective planes which contain only finitely many points. We have already noted that for each prime p there is a corresponding projective plane $\mathbb{Z}_p\mathbb{P}^2$. As noted below, these are more than abstract curiosities, and they play an important role in combinatorial theory (the study of counting principles) and its applications to experimental design and error-correcting codes.

By Theorem 9, it follows that a projective plane is finite if and only if its dual plane is finite. In fact, one can draw much stronger conclusions.

THEOREM IV.11. *Let P be a finite projective plane. Then all lines in P contain exactly the same number of points.*

Proof. Let L and M be the lines. Since there exist four point, no three of which are collinear, there must exist a point $\mathbf{p}$ which belongs to neither L nor M.

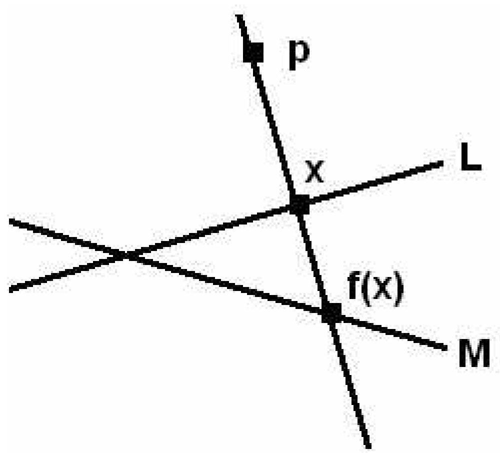


Figure IV.8

Define a map $f : L \to M$ by sending $\mathbf{x} \in L$ to the point $f(\mathbf{x})$ where $\mathbf{px}$ meets M. It is a straightforward exercise to verify that f is 1–1 and onto (see Exercise 6 below).■

Dualizing the preceding, we obtain the following conclusion.

THEOREM IV.12. *Let P be a finite projective plane. Then all points in P are contained in exactly the same number of lines.*■

Observe that the next result is self-dual, with the dual statement logically equivalent to the original one.

THEOREM IV.13. *Let P be a finite projective plane. Then the number of points on each line is equal to the number of lines containing each point.*

Proof. Let L be a line in P and let $\mathbf{x} \notin L$. Define a map from lines through $\mathbf{x}$ to points on L by sending a line M with $\mathbf{x} \in M$ to its unique intersection point with L. It is again a routine exercise to show this map is 1–1 and onto (again see Exercise 6 below).■

Definition. The *order* of a finite projective plane is the positive integer $n \geq 2$ such that every line contains $n+1$ points and every point lies on $n+1$ lines. The reason for the subtracting one from the common number is as follows: If $\mathbb{F}$ is a finite field with q elements, then the order of $\mathbb{FP}^2$ will be equal to q.

The results above yield the following interesting and significant restriction on the number of points in a finite projective plane:

THEOREM IV.14. *Let P be a finite projective plane of order n. Then P contains exactly n^2+n+1 points.*

In particular, for most positive integers m it is not possible to construct a projective plane with exactly m points. More will be said about the possibilities for m below.

Proof. We know that $n+1$ is the number or points on every line and the number of lines through every point. Let $\mathbf{x} \in P$. If we count all pairs $(\mathbf{y}, L)$ such that L is a line through $\mathbf{x}$ and $\mathbf{y} \in L$, then we see that there are exactly $(n+1)^2$ of them. In counting the pairs, some points such as $\mathbf{x}$ may appear more than once. However, $\mathbf{x}$ is the only point which does so, for $\mathbf{y} \neq \mathbf{x}$ implies there is only one line containing both points. Furthermore, by the preceding result we know that $\mathbf{x}$ appears exactly $n+1$ times. Therefore the correct number of points in P is given by subtracting n (not $n+1$) from the number of ordered pairs, and it follows that P contains exactly

$$(n+1)^2 \; - \; n \; = \; n^2 \; + \; n \; + \; 1$$

distinct points.■

The next theorem follows immediately by duality.

THEOREM IV.15. *Let P be a finite projective plane of order n. Then P contains exactly n^2+n+1 lines.*■

Further remarks on finite projective planes

We shall now consider two issues raised in the preceding discussion:

(1) The possible orders of finite projective planes.

(2) The mathematical and nonmathematical uses of finite projective planes.

ORDERS OF FINITE PROJECTIVE PLANES. The theory of finite fields is completely understood and is presented in nearly every graduate level algebra textbook (for example, see Section V.5 of the book by Hungerford in the bibliography). For our purposes it will suffice to note that for each prime number p and each positive integer n, there is a field with exactly $q = p^n$ elements. It follows that *every prime power is the order of some projective plane.*

The possible existence of projective planes with other orders is an open question. However, many possible orders are excluded by the following result, which is known as the **Bruck–Ryser–Chowla Theorem:** *Suppose that the positive integer n has the form $4k+1$ or $4k+2$ for some positive integer k. Then n is the sum of two (integral) squares.*

References for this result include the books by Albert and Sandler, Hall (the book on group theory), and Ryser in the bibliography as well as the original paper by Bruck-Ryser (also in the bibliography) and the following online reference[6]:

`http://www.math.unh.edu/~dvf/532/7proj-plane.pdf`

Here is a list of the integers between 2 and 100 which cannot be orders of finite projective planes by the Bruck-Ryser-Chowla Theorem:

6	14	21	22	30	33	38	42
45	46	54	57	62	66	69	70
74	77	78	82	86	93	94	97

The smallest positive integer ≥ 2 which is not a prime power and not excluded by the Bruck-Ryser-Chowla Theorem is 10. In the late nineteen eighties a substantial argument — which used sophisticated methods together with involved massive amounts of computer calculations — showed that no projective planes of order 10 exist. Two papers on this work are listed in the bibliography; the article by one author (C. W. H. Lam) in the *American Mathematical Monthly* was written to explain the research on this problem to a reasonably broad general audience of mathematicians and students.

By the preceding discussion, the existence of projective planes of order n is understood for $n \leq 11$, and the first open case is $n = 12$. Very little is known about this case.

PROJECTIVE GEOMETRY AND FINITE CONFIGURATIONS. Of course, one can view existence problems about finite projective planes as extremely challenging puzzles similar to Magic Squares (including the *Sudoku* puzzles that have recently become extremely popular), but one important reason for studying them is their relevance to questions of independent interest.

One somewhat whimsical "application" (*plant ten trees in ten rows of three*) was mentioned in Exercise IV.2.2, where the point was that such configurations exist by Desargues' Theorem. In fact, the study of projective spaces — especially finite ones – turns out to have many useful consequences in the study of finite *tactical configurations* or *block designs*, which is part of *combinatorial theory* or *combinatorics*. Specifically, finite projective planes are of interest as examples of *Latin squares*, or square matrices whose entries are in a finite set such that each element appears in every row and every column exactly once. As noted earlier, such objects play a significant role in the areas of statistics involving the design of experiments and in the theory of error-correcting codes. Although further comments are well outside the scope of these notes, the references in the bibliography by Buekenhout, Crapo and Rota, Hall (the book on combinatorial theory), Kárteszi, Lindner and Rodgers, and (the previously cited book by) Ryser are all sources for further information. There are also several online web sites dedicated to questions about finite geometry.

[6]To update this document, the conjecture of E. Catalan (1814–1894) has recently been shown to be true bu P. Mihăilescu. The proof is at a very advanced level, but for the sake of completeness here is a reference: P. Mihăilescu, *Primary cyclotomic units and a proof of Catalan's Conjecture*, [Crelle] Journal für die reine und angewandte Mathematik **572** (2004), 167–195.

Duality in higher dimensions

The concept of duality extends to all projective n-spaces (where $n \geq 2$), but the duality is more complicated for $n \geq 3$ than it is in the 2-dimensional case. Specifically, if S is a projective n-space, then points of the dual space S^* are given by the *hyperplanes* of S. for each k-plane P in S, there is an associated *(linear) bundle of hyperplanes with center* P, which we shall denote by $\mathfrak{b}(P)$, and the dimension of $\mathfrak{b}(P)$ is set equal to $n - k - 1$. We shall also denote the set of all such bundles by Π^* the associated dimension function by d^*. The following result is the appropriate generalization of Theorems 6 and 7 which allows extension of the principle of duality to higher dimensions.

THEOREM IV.16. *If (S, Π, d) is a projective n-space for some $n \geq 2$, then so is the dual object (S^*, Π^*, d^*). Furthermore, the map $\mathcal{E}$ sending $\mathbf{x} \in S$ to the bundle of hyperplanes $\mathfrak{b}(\mathbf{x})$ with center x defines an isomorphism of geometrical incidence spaces.*

We shall not give a direct proof of this result for two reasons.

1. Although the proof is totally elementary, it is a rather long and boring sequence of routine verifications.

2. The result follows from a coordinatization theorem in the next section (Theorem 18) and the results of Section VI.1, at least if $n \geq 3$ (and we have already done the case $n = 2$).

An excellent direct proof of Theorem 16 is given in Sections 4.3 and 4.4 of Murtha and Willard, *Linear Algebra and Geometry* (see the bibliography for further information).

EXERCISES

1. Prove that properties $(1) - (1^*)$ and $(2) - (2^*)$ completely characterize projective planes. In other words, if $(P, \mathcal{L})$ is a pair consisting of a set P and a nonempty collection of proper subsets $\mathcal{L}$ satisfying these, then there is a geometrical incidence space structure $(P, \mathcal{L}, d)$ such that the incidence space is a projective plane and $\mathcal{L}$ is the family of lines in Π.

2. Write out the plane duals of the following phrases, and sketch both the given data and their plane duals.

(i) Two lines, and a point on neither line.

(ii) Three collinear points and a fourth point not on the line of the other three.

(iii) Two triples of collinear points not on the same line.

(iv) Three nonconcurrent lines, and three points such that each point lies on exactly one of the lines.

3. Draw the plane duals of the illustrated finite sets of points which are marked heavily in the two drawings below.

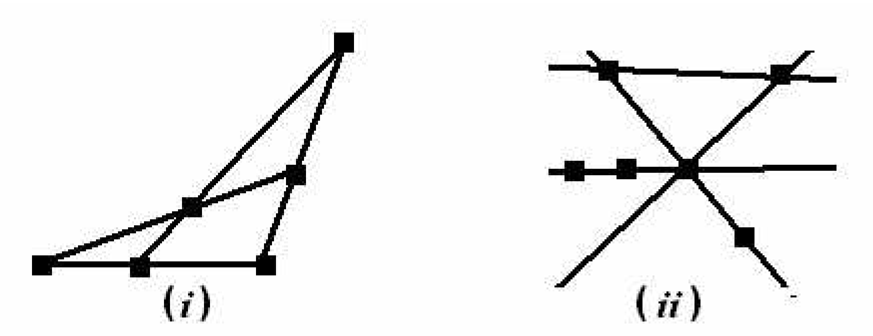


Figure IV.9

4. Suppose that $f : S \to T$ is an incidence space isomorphism from one projective n-space ($n \geq 2$) to another. Prove that f induces a 1–1 correspondence $f^* : S^* \to T^*$ taking a hyperplane $H \subset S$ to the image hyperplane $f^*(H) \;=\; f[H] \;\subset T$. Prove that this correspondence has the property that B is an r-dimensional bundle of hyperplanes (in S^*) if and only if $f^*[B]$ is such a subset of T^* (hence it is an isomorphism of geometrical incidence spaces, assuming Theorem 16).

5. Prove that the construction in the preceding exercise sending f to f^* has the following properties:

(i) If $g : T \to U$ is another isomorphism of projective n-spaces, then $(g \circ f)^* \;=\; g^* \circ f^*$.

(ii) For all choices of S the map $(\mathrm{id}_S)^*$ is equal to the identity on S^*.

(iii) For all f we have $\left(f^{-1}\right)^* \;=\; \left(f^*\right)^{-1}$.

6. Complete the proof of Theorem 11.

7. Let $(P, \mathcal{L})$ be a finite affine plane. Prove that there is a positive integer n such that

(i) every line in P contains exactly n points,

(ii) every point in P lies on exactly n lines,

(iii) the plane P contains exactly n^2 points.

4. Conditions for coordinatization

Since the results and techniques of linear algebra are applicable to coordinate projective n-spaces $\mathcal{S}_1(V)$ (where $\dim V = n+1$), these are the most conveniently studied of all projective spaces. Thus it is desirable to know when a projective n-space is isomorphic to one having the form $\mathcal{S}_1(V)$, where $\dim V = n+1$. The following remarkable theorem shows that relatively weak hypotheses suffice for the existence of such an isomorphism.

THEOREM IV.17. *Let P be a projective n-space in which Desargues' Theorem is valid (for example, $n \geq 3$ or P is a Desarguian plane). Then there is a skew-field $\mathbb{F}$ such that P is isomorphic to $\mathbb{FP}^n$ (where we view $\mathbb{F}^{n+1}$ as a right vector space over $\mathbb{F}$. If $\mathbb{E}$ is another skew-field such that P is isomorphic to $\mathbb{EP}^n$, then $\mathbb{E}$ and $\mathbb{F}$ are isomorphic as skwe-fields.*

A well-illustrated proof of Theorem 17 from first principles in the case $n = 2$ appears on pages 175–193 of the book by Fishback listed in the bibliography. Other versions of the proof appear in several other references from the bibliography. Very abstract approaches to the theorem when $n = 2$ appear in Chapter III of the book by Bumcrot and also in the book by Artzy. The proof in Chapter 6 of Hartshorne's book combines some of the best features of the other proofs. There is also a proof of Theorem 17 for arbitrary values of $n \geq 2$ in Chapter VI from Volume I of Hodge and Pedoe. Yet another reference is Sections 4.6 and 4.7 of Murtha and Willard. For more information on the uniqueness statement, see Theorem V.10.

Theorem 17 yields a classification for projective n-spaces ($n \geq 3$) that is parallel to Theorem II.38 (see Remark 3 below for further discussion). Because of its importance, we state this classification separately.

THEOREM IV.18. *Let P be a projective n-space, where $n \geq 3$. Then there is a skew-field $\mathbb{F}$, unique up to algebraic isomorphism, such that P and $\mathbb{FP}^n$ are isomorphic as geometrical incidence spaces.*■

Theorem 17 also implies the converse to a remark following the definition of a Desarguian projective plane in Section IV.2.

THEOREM IV.19. *If a projective plane is Desarguian, then it is isomorphic to a plane in a projective 3-space.*

Proof. By Theorem 17, the plane is isomorphic to $\mathbb{FP}^2$ for some skew-field $\mathbb{F}$. But $\mathbb{FP}^2$ is isomorphic to the plane in $\mathbb{FP}^3$ defined by $x_4 = 0$.■

REMARK 1. Suppose that P is the Desarguian plane $\mathbb{FP}^2$. By Theorem 10, P^* is also Desarguian and hence is isomorphic to some plane $\mathbb{EP}^2$. As one might expect, the skew-fields $\mathbb{F}$ and $\mathbb{E}$ are closely related. In fact, by Theorem V.1 the dual plane P^* is isomorphic to $\mathcal{S}_1(\mathbb{F}_3)$, where $\mathbb{F}_3$ is a *left* vector space over $\mathbb{F}$, and left vector spaces over $\mathbb{F}$ correespond to right vector spaces over the **opposite skew-field** $\mathbb{F}^{\mathsf{OP}}$, whose elements are the same as $\mathbb{F}$ and whose multiplication is given by reversing the multiplication in $\mathbb{F}$. More precisely, one defines a new product $\otimes$ in $\mathbb{F}$ via $a \otimes b = b \cdot a$, and if V is a left vector space over $\mathbb{F}$ define vector space operations via the vector addition on $\mathbb{F}$ and the right scalar product $\mathbf{x} \otimes a \;=\; a \cdot \mathbf{x}$ (here the dot represents the original multiplication). It is a routine exercise to check that $\mathbb{F}^{\mathsf{OP}}$ is a skew-field and $\otimes$ makes left $\mathbb{F}$-vector spaces into right vector spaces over $\mathbb{F}^{\mathsf{OP}}$. Thus, since $\mathbb{F}_3$ is a 3-dimensional right $\mathbb{F}^{\mathsf{OP}}$-vector space, it follows that $\mathbb{E}$ must be isomorphic to $\mathbb{F}^{\mathsf{OP}}$.

REMARK 2. If P is the coordinate projective plane $\mathbb{FP}^2$ and multiplication in $\mathbb{F}$ is commutative, then the preceding remark implies that P and P^* are isomorphic because $\mathbb{F}$ and $\mathbb{F}^{\mathsf{OP}}$ are identical in such cases. However, the reader should be warned that the isomorphisms from P to P^* are much less "natural" than the standard isomorphisms from P to P^{**} (this is illustrated by Exercise V.1.5). More information on the noncommutative case appears in Appendix C.

EXAMPLE 3. Theorem II.18 follows from Theorem 18. For if S is an affine n-space, let $\overline{S}$ be its synthetic projective extension as defined in the Addendum to Section III.4. By Theorem III.16, we know that $\overline{S}$ is an n-dimensional projective incidence space, so that $\overline{S}$ is isomorphic to $\mathbb{FP}^n$ for some skew-field $\mathbb{F}$ and $S \subset \overline{S}$ is the complement of some hyperplane. Since there is an element of the geometric symmetry group of $\mathbb{FP}^n$ taking this hyperplane to the one defined by $x_{n+1} = 0$ (see Exercise III.4.14), we can assume that S corresponds to the image of $\mathbb{F}^n$ in $\mathbb{FP}^n$. Since k-planes in S are given by intersections of k-planes in $\overline{S}$ with S and similarly the k-planes in $\mathbb{F}^n$ are given by intersections of k-planes in $\mathbb{FP}^n$ with the image of $\mathbb{F}$, it follows that the induced 1–1 correspondence between S and $\mathbb{F}^n$ is a geometrical incidence space isomorphism.

A general coordinatization theory for projective planes that are not necessarily Desarguian exists; the corresponding algebraic systems are generalizations of skew-fields known as *planar ternary rings*. Among the more accessible references for this material are the previously cited book by Albert and Sandler, Chapter 4 of the book by Artzy, Chapters III and VI of the book by Bumcrot, Chapter 17 of the book by Hall on group theory, and the 2007 survey article by C. Weibel (see the bibliography for more details).

One of the most important examples of a planar ternary ring (the **Cayley numbers**) is described on pages 195–196 of Artzy[7] As one might expect, the Cayley numbers (also called *octonions*) are associated to a projective plane called the *Cayley projective plane*. However, the coordinatization theorem implies that are no Cayley projective spaces of higher dimension.

[7]One brief and accurate online summary is given in `http://en.wikipedia.org/wiki/Octonion`. The reliability issue for Wikipedia articles is discussed in a footnote for Appendix C, and it also applies here.

CHAPTER V

PLANE PROJECTIVE GEOMETRY

In this chapter we shall present the classical results of plane projective geometry. For the most part, we shall be working with coordinate projective planes and using homogeneous coordinates, but at certain points we shall also use synthetic methods, especially when it is more convenient to do so. Our treatment will make extensive use of concepts from linear algebra. Since one major geometric result (Pappus' Theorem) is closely connected to the algebraic commutativity of multiplication in a skew-field, we shall be fairly specific about using left or right vector spaces in most sections of this chapter.

1. Homogeneous line coordinates

If $\mathbb{F}$ is a skew-field, it will be convenient to let view $\mathbb{FP}^n$ as the set of all 1-dimensional vector subspaces of the $(n+1)$-dimensional right vector space $\mathbb{F}^{n+1,1}$ of $(n+1) \times 1$ column matrices over $\mathbb{F}$ with the obvious entrywise right multiplication:

$$\begin{pmatrix} x_1 \\ \cdots \\ \cdots \\ x_{n+1} \end{pmatrix} \cdot c \;=\; \begin{pmatrix} x_1\, c \\ \cdots \\ \cdots \\ x_{n+1}\, c \end{pmatrix}$$

It follows from Theorem III.12 that a line in $\mathbb{FP}^2$ is definable by an equation of the form $u_1x_1 + u_2x_2 + u_3x_3 = 0$, where $u_1,\ u_2,\ u_3$ are not all zero. Furthermore, two triples of coefficients $(u_1,\ u_2,\ u_3)$ and $(v_1,\ v_2,\ v_3)$ define the same line if and only if there is a nonzero $k \in \mathbb{F}$ such that $U_i = k\, v_i$ for $i = 1, 2, 3$. Thus we see that *a line in $\mathbb{FP}^2$ is completely determined by a one-dimensional subspace of the* left *vector space of 1×3 row matrices.* — Therefore the dual projective plane to $\mathbb{FP}^2$ is in 1–1 correspondence with the 1-dimensional subspaces of $\mathbb{F}^{1,3}$, where the latter is the left vector space of 1×3 matrices. Under this correspondence the lines in the dual of $\mathbb{FP}^2$ correspond to lines in $\mathcal{S}_1(\mathbb{F}^{1,3})$. For a line in the dual is the set of lines through a given point, and by a reversal of the previous argument a line in $\mathcal{S}_1(\mathbb{F}^{1,3})$ is just the set of elements whose homogeneous coordinates $(u_1,\ u_2,\ u_3)$ satisfy a linear homogeneous equation of the form $u_1x_1 + u_2x_2 + u_3x_3 = 0$, where $x_1,\ x_2,\ x_3$ are not all zero. Much as before, the triples $(x_1,\ x_2,\ x_3)$ and $(y_1,\ y_2,\ y_3)$ define the same set if and only if $y_i = x_i\, m$ for some nonzero constant $m \in \mathbb{F}$. We can summarize the discussion above as follows:

THEOREM V.1. *Let $\mathbb{F}$ be a skew-field, and identify $\mathbb{FP}^2$ with $\mathcal{S}_1(\mathbb{F}^{3,1})$ as above. Then the dual plane $\left(\mathbb{FP}^2\right)^*$ is isomorphic to $\mathcal{S}_1(\mathbb{F}^{1,3})$ such that if the point $\mathbf{x}$ corresponds to the 1-dimensional right vector subspace $\xi \cdot \mathbb{F}$ and the line L corresponds to the 1-dimensional left vector subspace $\mathbb{F} \cdot \lambda$, then $\mathbf{x} \in L$ if and only if $\lambda \cdot \xi = 0$.* ■

The dot indicates matrix multiplication operation

$$\mathbb{F}^{1,3} \times \mathbb{F}^{3,1} \longrightarrow \mathbb{F}^{1,1} \cong \mathbb{F} .$$

As before, $\mathbb{F} \cdot \lambda$ denotes all left scalar multiples of λ, and we similarly let $\xi \cdot \mathbb{F}$ denote all right scalar multiples of ξ. Note that if $\lambda \cdot \xi = 0$ for one pair of homogeneous coordinate choices, then the same is true for every other pair. For the most general change of representatives is given by $k\,\lambda$ and $\xi\,m$ so that

$$(k\,\lambda) \cdot (\xi\,m) \;=\; k\,(\lambda \cdot \xi)\,m \;=\; 0$$

by associativity of multiplication.

Definition. Given a line L in $\left(\mathbb{FP}^2\right)^* \cong \mathcal{S}_1(\mathbb{F}^{1,3})$, we say that a nonzero vector $\lambda \in \mathbb{F}^{1,3}$ is a set of homogeneous coordinates for L if the latter is the set of all points $\mathbf{x}$ whose homogeneous coordinates ξ satisfy $\lambda \cdot \xi = 0$.

By construction, three points in $\mathbb{FP}^2$ are collinear if and only if their homogeneous coordinates span a 2-dimensional right vector subspace of $\mathbb{F}^{3,1}$. The dualization of this to homogeneous line coordinates is an easy consequence of Theorem 1.

THEOREM V.2. *Let L, M and N be three distinct lines in $\mathbb{FP}^2$. Then they are concurrent if and only if their homogeneous line coordinates are linearly dependent.*$\blacksquare$

NOTATIONAL CONVENTIONS. Throughout this chapter we shall be passing back and forth between geometric points and lines and the algebraic homogeneous coordinates which represent them. Needless to say, it is convenient to have some standard guidelines for passing back and forth between the geometric and algebraic objects. Normally we shall denote the geometric objects by Roman letters and appropriate homogeneous coordinates by corresponding Greek letters (strictly speaking, we use mathematicians' versions of Greek letters). For example, if X and Y are points, then we shall normally use ξ or η for homogeneous coordinates, and if L is a line we shall normally use λ.

Homogeneous coordinate formulas

In the remainder of this section, we shall describe some useful formulas which are valid provided the skew-field $\mathbb{F}$ is **commutative**. Of course, if $\mathbb{F}$ is commutative the distinction between left and right vector subspaces is unnecessary.

We begin by stating two obvious problems:

1. If L is a line determined by $\mathbf{x}$ and $\mathbf{y}$, express homogeneous coordinates for L in terms of homogeneous coordinates for $\mathbf{x}$ and $\mathbf{y}$.
2. If $\mathbf{x}$ is the point of intersection for lines L and M, express homogeneous coordinates for $\mathbf{x}$ in terms of homogeneous coordinates for L and M.

Consider the first problem. By the definition of homogeneous coordinates of lines, a set of homogeneous coordinates λ for L must satisfy $\lambda \cdot \xi = \lambda \cdot \eta = 0$. If $\mathbb{F}$ is the real numbers, this means that the transpose of λ is perpendicular to ξ and η. Since $\mathbf{x} \neq \mathbf{y}$, it follows that ξ and η are linearly independent; consequently, the subspace of vectors perpendicular to both of the latter must be 1-dimensional. A nonzero (and hence spanning) vector in the subspace of vectors perpendicular to ξ and η is given by the *cross product* $\xi \times \eta$, where the latter are viewed as ordinary 3-dimensional vectors (see Section 5 of the Appendix). It follows that λ may be chosen to be an arbitrary nonzero multiple of $\xi \times \eta$. We shall generalize this formula to other fields.

THEOREM V.3. *Let $\mathbb{F}$ be a (commutative) field, and let $\mathbf{x}$ and $\mathbf{y}$ be distinct points in $\mathbb{FP}^2$ having homogeneous coordinates ξ and η. Then the line $\mathbf{xy}$ has homogeneous coordinates given by the* transpose *of $\xi \times \eta$.*

Proof. The definition of cross product implies that

$$ {}^{\mathbf{T}}(\xi \times \eta) \cdot \xi \;=\; {}^{\mathbf{T}}(\xi \times \eta) \cdot \eta \;=\; 0 $$

so it is only necessary to show that if ξ and η are linearly independent then $\xi \times \eta \neq \mathbf{0}$.

Let the entries of ξ be given by x_i, let the entries of η be given by y_j, and consider the 3×2 matrix B whose entries are the entries of the 3×1 matrices ξ and η:

$$ \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} $$

Since the columns are linearly independent, the rank of this matrix is 2. By Theorem A.11 this means there is 2×2 submatrix of B with nonzero determinant. If $k \in \{1, 2, 3\}$ is such that the matrix obtained by deleting the k^{th} row is nonzero, then by the definition of cross product the k^{th} entry of the latter must be nonzero. Therefore the transpose of $\xi \times \eta$ is a set of homogeneous coordinates for L.$\blacksquare$

Dually, we have the following result:

THEOREM V.4. *Let $\mathbb{F}$ be a (commutative) field, and let L and M be distinct lines in $\mathbb{FP}^2$ having homogeneous coordinates λ and μ. Then the intersection point of L and M has homogeneous coordinates given by the* transpose *of $\lambda \times \mu$.*$\blacksquare$

The "Back-Cab Rule" for triple cross products

$$ \mathbf{a} \;\times\; (\mathbf{b} \times \mathbf{c}) \;=\; \mathbf{b}\,(\mathbf{a} \cdot \mathbf{c}) \;-\; \mathbf{c}\,(\mathbf{a} \cdot \mathbf{b}) $$

(see Theorem A.20) implies the following useful formula:

THEOREM V.5. *Let L be a line in $\mathbb{FP}^2$, and let $\mathbf{x}$ and $\mathbf{y}$ be points of $\mathbb{FP}^2$ not on L. Let λ, ξ and η be homogeneous coordinates for L, $\mathbf{x}$ and $\mathbf{y}$ respectively. Then the common point of the lines L and $\mathbf{xy}$ has homogeneous coordinates equal to $(\lambda \cdot \xi)\,{}^{\mathbf{T}}\eta \;-\; (\lambda \cdot \eta)\,{}^{\mathbf{T}}\xi$.*

Proof. By Theorem 3 the line $\mathbf{xy}$ has homogeneous coordinates ${}^{\mathbf{T}}\xi \times {}^{\mathbf{T}}\eta$, and thus by Theorem 4 the common point of L and $\mathbf{xy}$ has homogeneous coordinates equal to

$$ {}^{\mathbf{T}}\big(\, \lambda \times {}^{\mathbf{T}}\big({}^{\mathbf{T}}\xi \times {}^{\mathbf{T}}\eta\big)\,\big) \ . $$

The latter is equal to $(\lambda \cdot \xi)\,{}^{\mathbf{T}}\eta \ - \ (\lambda \cdot \eta)\,{}^{\mathbf{T}}\xi$ by the Back-Cab Rule.■

EXERCISES

1. Consider the affine line in $\mathbb{F}^2$ defined by the equation $ax + by \ = \ c$. What are the homogeneous coordinates of its extension to $\mathbb{FP}^2$? As usual, consider the 1–1 map from $\mathbb{F}^2$ to $\mathbb{FP}^2$ which sends $(x,\, y)$ to the point with homogeneous coordinates

$$ \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \ . $$

2. Suppose that three affine lines are defined by the equations $a_i x + b_i y \ = \ c_i$, where $i = 1, 2, 3$. Prove that these three lines are concurrent if and only if

$$ \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \ = \ 0 \qquad \text{and} \qquad \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \ \neq \ 0 \ . $$

What can one conclude about the lines if **both** determinants vanish?

3. Using the methods of this section, find the equation of the affine line joining $(\frac{1}{4},\, \frac{1}{2})$ to the point of intersection of the lines defined by the equations $x + 2y + 1 \ = \ 0$ and $2x + y + 3 \ = \ 0$.

4. Fine the homogeneous coordinates of the point at which the line with homogeneous coordinates $(2 \quad 1 \quad 4)$ meets the line through the points with homogeneous coordinates

$$ \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \ . $$

5. Let A be an invertible 3×3 matrix over $\mathbb{F}$, and let T be the geometric symmetry of $\mathbb{FP}^2 \ \cong \ \mathcal{S}_1(\mathbb{F}^{3,1})$ defined by the equation

$$ T(\mathbf{x}) \ = \ A \cdot \xi \cdot \mathbb{F} $$

where ξ is a set of homogeneous coordinates for $\mathbf{x}$ and the dot indicates matrix multiplication. If L is a line in $\mathbb{FP}^2$ with homogeneous coordinates λ, show that the line $f[L]$ has homogeneous coordinates given by $\lambda \cdot A^{-1}$.

2. Cross ratio

At the beginning of Chapter III, we mentioned that one forerunner of projective geometry was the development of a mathematical theory of perspective images by artists during the 15^{th} and early 16^{th} century. Clearly, if one compares such perspective photographic images with the physical objects they come from, it appears that some physical properties are faithfully captured by the photograph while others are not. For example, if three points on a physical object are collinear, then their photographic images are also collinear, and under suitable conditions if we have three distinct points A, B and C such that B is between A and C, then the image point B' of B will also lie between the corresponding image points A and C'.[1] However, it is also apparent that the relative distances among the three points can be greatly distorted. For example, if B is the midpoint of A and C, then we cannot conclude that B' is the midpoint of A' and C'. Similarly, if B is between A and C and the distance from B to C is twice the distance from A to B, we cannot conclude that a similar relationship holds for the corresponding relative distances among the image points A', B' and C'. HOWEVER, if we are given *four* collinear points on the physical object, then there is a number called the *cross ratio*, which is determined by relative distances among the four points, that is the same for the original four points on the physical object(s) as well as their photographic images. The cross ratio itself was apparently first defined algebraically by P. de la Hire (1640–1718), but the perspective invariance property was essentially known to Pappus of Alexandria (*c.* 290 – *c.* 350??) and perhaps even earlier. Throughout the rest of these notes we shall see that the cross ratio plays a fundamentally important role in projective geometry.

It will be convenient to define the cross ratio in terms of homogeneous coordinates and to give a nonvisual motivation for the concept. In problems involving coordinate projective spaces, it is often helpful to choose homogeneous coordinates in a particular way. We shall prove some results justifying such choices below and use them to give a fairly simple definition of the cross product. The discussion up to (but not including) Theorem 8 is valid for any skew-field $\mathbb{F}$. Starting with Theorem 9, we assume $\mathbb{F}$ is commutative.

THEOREM V.6. *Let A, B, C, D be four points in $\mathbb{FP}^2$, no three of which are collinear. Then there exist homogeneous coordinates α, β, γ, δ for A, B, C, D such that $\delta = \alpha + \beta + \gamma$.*

Proof. Let α_0, β_0, γ_0 be arbitrary homogeneous coordinates for A, B, C respectively. Since A, B, C are noncollinear, the vectors α_0, β_0, γ_0 form a basis for $\mathbb{F}^{3,1}$. Thus there exist $x, y, z \in \mathbb{F}$ such that homogeneous coordinates for D are given by $\delta = \alpha_0 x + \beta_0 y + \gamma_0 z$. We claim that each of x, y, z is nonzero. If $x = 0$, then it follows that $\delta = y\beta_0 + z\gamma_0$, so that D lies on the line BC; therefore it follows that $x \neq 0$, and similarly we can conclude that y and z are nonzero. But this means that $\alpha_0 x$, $\beta_0 y$, $\gamma_0 z$ are homogeneous coordinates for A, B, C respectively, and accordingly we may take $\alpha = \alpha_0 x$, $\beta = \beta_0 y$, and $\gamma = \gamma_0 z$.■

The next few results are true in $\mathbb{FP}^n$ for any $n \geq 1$.

[1] The suitable, and physically reasonable, conditions are given in terms of the aperture of the camera at some point X and the image plane P which does not contain X: If Q is the unique plane through X which is parallel to P, then normally the plane P the physical object(s) being photographed will lie on opposite sides of the plane Q.

THEOREM V.7. *Let A and B Let A and B be distinct points in $\mathbb{FP}^n$ with homogeneous coordinates α and β respectively. Let C be a third point on AB. Then there exist homogeneous coordinates γ for C such that $\gamma = \alpha x + \beta$ for some unique $x \in \mathbb{F}$.*

Proof. Since A, B, and C are noncollinear, there exist u, $v \in \mathbb{F}$ such that homogeneous coordinates for C are given by $\alpha u + \beta v$. We claim that $v \neq 0$, for otherwise $\gamma = \alpha u$ would imply $C = A$. Thus $\alpha u\, v^{-1} + \beta$ is also a set of homogeneous coordinates for C, proving the existence portion of the theorem. Conversely, if $\alpha y + \beta$ is also a set of homogeneous coordinates for C, then there is a nonzero scalar k such that

$$\alpha y \;+\; \beta \;=\; (\alpha x \;+\; \beta) \cdot k \;=\; \alpha x k \;+\; \beta k \;.$$

Equating coefficients, we have $k = 1$ and $y = xk = x$.■

Notation. If $C \neq A$, the element of $\mathbb{F}$ determined by Theorem 7 is called the *nonhomogeneous coordinate* of C with respect to α and β and written $\gamma_{(\alpha,\beta)}$.

THEOREM V.8. *Let A, B, and C be distinct collinear points in $\mathbb{FP}^n$. Then there exist homogeneous coordinates α, β, γ for A, B, C such that $\gamma = \alpha + \beta$. Furthermore, if α', β', γ' are arbitrary homogeneous coordinates for A, B, and C, then there is a nonzero constant $k \in \mathbb{F}$ such that $\alpha' = \alpha k$, $\beta' = \beta k$, and $\gamma' = \gamma k$.*

Proof. By Theorem 7 there exist homogeneous coordinates α'' and β'' for A and B such that homogeneous coordinates for C are given by $\gamma = \alpha'' x + \beta''$. If x were zero then B and A would be equal, and consequently we must have $x \neq 0$. Thus if we take $\alpha = \alpha'' x$ and $\beta = \beta'$, then $\gamma = \alpha + \beta$ is immediate.

Suppose that α', β', γ' satisfy the condition of the theorem. Then there exist constants a, b, c such that $\alpha' = \alpha\, a$, $\beta' = \beta\, b$, and $\gamma' = \gamma\, c$. It follows that

$$\alpha c \;+\; \beta c \;=\; \gamma c \;=\; \alpha' \;+\; \beta' \;=\; \alpha \cdot a \;+\; \beta \cdot b \;.$$

Equating coefficients, we obtain $a \;=\; c \;=\; b$, and thus we may take $k \;=\; c$.■

COMMUTATIVITY ASSUMPTION. Throughout the rest of this section $\mathbb{F}$ is assumed to be commutative. The definition of cross ratio is justified by the following result:

THEOREM V.9. *Let A, B and C be distinct collinear points in $\mathbb{FP}^n$, and let D be a point on this line such with $D \neq A$. Suppose that homogeneous coordinates α, β, γ for A, B, C are chosen such that $\gamma = \alpha + \beta$, and write homogeneous coordinates for D as $\delta = u\alpha + v\beta$ in these coordinates (since $D \neq A$ we must have $v \neq 0$). Then the quotient u/v is the same for all choices of α, β, γ satisfying the given equation.*

Definition. The scalar $u/v \in \mathbb{F}$ is called the **cross ratio** of the ordered quadruple of collinear points (A, B, C, D), and it is denoted by $\mathsf{XR}(A, B, C, D)$.
Proof. If α', β', γ' is another triple such that $\gamma' = \alpha' + \beta'$, then by Theorem 8 there is a nonzero scalar $k \in \mathbb{F}$ such that $\alpha' = k\alpha$, $\beta' = k\beta$, and $\gamma' = k\gamma$. If δ' is another set of homogeneous coordinates for D, then $\delta' = r\delta$ for some $r \in \mathbb{F}$. Thus if $\delta' = u'\alpha' + v'\beta'$, it follows that $\delta' = ku'\alpha + kv'\beta$. On the other hand, $\delta' = r\delta$ implies that $\delta' = ru\alpha + rv\beta$. Equating coefficients, we find that $ku' = ru$ and $kv' = rv$. Therefore we have

$$\frac{u}{v} \;=\; \frac{ru}{rv} \;=\; \frac{ku'}{kv'} \;=\; \frac{u'}{v'}$$

and therefore the ratio of the coefficients does not depend upon the choices of homogeneous coordinates.■

The next result answers a fundamental question concerning the cross ratio:

THEOREM V.10. *Suppose that A, B and C are distinct collinear points and $r \in \mathbb{F}$ is an arbitrary constant. Then there is a unique $D \in AB$ such that* $\mathsf{XR}(A, B, C, D) \ = \ r$.

Proof. *Existence.* Choose D so that it has homogeneous coordinates equal to $\delta = r\alpha + \beta$.

Uniqueness. Suppose that $\mathsf{XR}(A, B, C, D) \ = \ \mathsf{XR}(A, B, C, E) \ = \ r$, where neither D nor E is equal to A. Choose homogeneous coordinates so that we have $\gamma = \alpha + \beta$; then we may write $\delta = u\alpha + v\beta$ and $\varepsilon = s\alpha + t\beta$, where

$$\frac{u}{v} \ = \ r \ = \ \frac{s}{t} \ ;$$

note that v and t are nonzero because neither D nor E is equal to A. It follows that $u = vr$ and $s = tr$, from which we conclude that $\delta = v\,t^{-1}\varepsilon$. The latter in turn implies that $D = E$.■

Another fundamental property of the cross ratio is given by the following result, whose proof is left as an exercise:

THEOREM V.11. *Let A, B, C and D be distinct noncollinear points, and let $E \neq A$ be another point on the same line. Then* $\mathsf{XR}(A, B, C, E) \ = \ \mathsf{XR}(A, B, C, D) \cdot \mathsf{XR}(A, B, D, E)$.■

There are 24 possible orders in which four distinct collinear points $A,\ B,\ C,\ D$ may be reordered. We summarize what happens to the cross ration under reordering below.

THEOREM V.12. *Let A, B, C and D be distinct noncollinear points in $\mathbb{FP}^n$, and assume that* $\mathsf{XR}(A, B, C, D) \ = \ r$. *Then the other* 23 *cross ratios involving these points by reordering are given as follows:*

$$\begin{aligned}
r \ &= \ \mathsf{XR}(A, B, C, D) \ = \ \mathsf{XR}(B, A, D, C) \ = \ \mathsf{XR}(C, D, A, B) \ = \ \mathsf{XR}(D, C, B, A) \\
\frac{1}{r} \ &= \ \mathsf{XR}(A, B, D, C) \ = \ \mathsf{XR}(B, A, C, D) \ = \ \mathsf{XR}(D, C, A, B) \ = \ \mathsf{XR}(C, D, B, A) \\
1 \ - \ r \ &= \ \mathsf{XR}(A, C, B, D) \ = \ \mathsf{XR}(C, A, B, D) \ = \ \mathsf{XR}(B, D, A, C) \ = \ \mathsf{XR}(D, B, C, A) \\
\frac{1}{1-r} \ &= \ \mathsf{XR}(A, C, D, B) \ = \ \mathsf{XR}(C, A, B, D) \ = \ \mathsf{XR}(D, B, A, C) \ = \ \mathsf{XR}(B, D, C, A) \\
\frac{1-r}{r} \ &= \ \mathsf{XR}(A, D, B, C) \ = \ \mathsf{XR}(D, A, C, B) \ = \ \mathsf{XR}(B, C, A, D) \ = \ \mathsf{XR}(C, B, D, A) \\
\frac{r}{1-r} \ &= \ \mathsf{XR}(A, D, C, B) \ = \ \mathsf{XR}(D, A, B, C) \ = \ \mathsf{XR}(C, B, A, D) \ = \ \mathsf{XR}(B, C, D, A)
\end{aligned}$$

The proof is a sequence of elementary and eventually boring calculations, and it is left as an exercise.■

The next result gives a useful expression for the cross ratio:

THEOREM V.13. *Let A_1, A_2, A_3 be distinct collinear points, and let $A_4 \neq A_1$ lie on this line. Let B_1, B_2, B_3 be distinct collinear points on this line with $B_1 \neq A_i$ for all i, and suppose that $\mathsf{XR}(B_1, B_2, B_3, A_i) = x_i$ for $i = 1, 2, 3, 4$. Then the following holds:*

$$\mathsf{XR}(A_1, A_2, A_3, A_4) \;=\; \frac{(x_1 - x_3)\,(x_2 - x_4)}{(x_1 - x_4)\,(x_2 - x_3)}$$

Proof. Choose homogeneous coordinates β_i for Q_i such that $\beta_3 = \beta_1 + \beta_2$. Then homogeneous coordinates α_j for the points A_j are given by $x_j\beta_1 + \beta_2$. It is not difficult to verify that

$$(x_1 - x_2)\alpha_3 \;=\; (x_3 - x_2)\alpha_1 \;+\; (x_1 - x_3)\alpha_2$$

is true and similarly

$$(x_1 - x_2)\alpha_3 \;=\; (x_3 - x_2)\alpha_1 \;+\; (x_1 - x_3)\alpha_2\ .$$

Thus if $\alpha'_1 = (x_3 - x_2)\beta_1$ and $\alpha'_2 = (x_1 - x_3)\beta_2$, then we have

$$(x_1 - x_2)\alpha_4 \;=\; \frac{(x_4 - x_2)}{(x_3 - x_2)}\,\alpha'_1 \;+\; \frac{(x_1 - x_4)}{(x_1 - x_3)}\,\alpha'_2\ .$$

The cross ratio formula in the theorem is an immediate consequence of these formulas.■

DUALIZATION. The preceding discussion can be dualized to yield the *cross ratio of four concurrent lines in* $\mathbb{FP}^2$. Cross rations of collinear points and concurrent lines are interrelated as follows.

THEOREM V.14. *Let L_1, L_2, L_3 be distinct concurrent lines, and let $L_4 \neq L_1$be another line through this point. Let M be a line in $\mathbb{FP}^2$ which does not contain this common point, and let A_i be the point where L_i meets M, where $i = 1, 2, 3, 4$. Then the point $A_4 \in M$ lies on L_4 if and only if* $\mathsf{XR}(A_1, A_2, A_3, A_4) \;=\; \mathsf{XR}(L_1, L_2, L_3, L_4)$.

Proof. *Suppose that $A_4 \in M \cap L_4$.* Let r be the cross ratio of the lines, and let s be the cross ratio of the points. Choose homogeneous coordinates for the points A_i and lines L_i such that $\alpha_3 = \alpha_1 + \alpha_2$ and $\lambda_3 = \lambda_1 + \lambda_2$, so that $\alpha_4 = s\,\alpha_1 + \alpha_2$ and $\lambda_4 = r\,\lambda_1 + \lambda_2$. Since $A_i \in L_i$ for all i, we have $\lambda_i \cdot \alpha_i = 0$ for all i. In particular, these equations also imply

$$0 \;=\; \lambda_3 \cdot \alpha_3 \;=\; (\lambda_1 + \lambda_2) \cdot (\alpha_1 + \alpha_2) \;=$$

$$\lambda_1 \cdot \alpha_1 \;+\; \lambda_1 \cdot \alpha_2 \;+\; \lambda_2 \cdot \alpha_1 \;+\; \lambda_2 \cdot \alpha_2 \;=$$

$$\lambda_1 \cdot \alpha_2 \;+\; \lambda_2 \cdot \alpha_1$$

so that $\lambda_1 \cdot \alpha_2 = -\lambda_2 \cdot \alpha_1$. Therefore we see that

$$0 \;=\; \lambda_r \cdot \alpha_r \;=\; (r\lambda_1 + \lambda_2) \cdot (s\alpha_1 + \alpha_2) \;=$$

$$r\lambda_1 \cdot \alpha_2 \;+\; s\lambda_2 \cdot \alpha_1 \;=\; (r - s)\lambda_1 \cdot \alpha_2\ .$$

The product $\lambda_1 \cdot \alpha_2$ is nonzero because $A_2 \notin L_1$, and consequently $r - s$ must be equal to zero, so that $r = s$.■

Suppose that the cross ratios are equal. Let $C \in M \cap L_4$; then by the previous discussion we know that

$$\mathsf{XR}(A_1, A_2, A_3, C) \;=\; \mathsf{XR}(L_1, L_2, L_3, L_4)\ .$$

Therefore we also have $\mathsf{XR}(A_1, A_2, A_3, C) \;=\; \mathsf{XR}(A_1, A_2, A_3, A_4)$, so that $A_4 = C$ by Theorem 10 and in addition we have $A_4 \in L_4$.■

The following consequence of Theorem 14 is the result on *perspective invariance of the cross ratio* mentioned at the beginning of this section.

THEOREM V.15. *Let L_1, L_2, L_3 and L_4 be distinct concurrent lines, and let M and N be two lines which do not contain the common point. Denote the intersection points of M and N with the lines L_i by A_i and B_i respectively. Then* $\mathsf{XR}(A_1, A_2, A_3, A_4) = \mathsf{XR}(B_1, B_2, B_3, B_4)$.

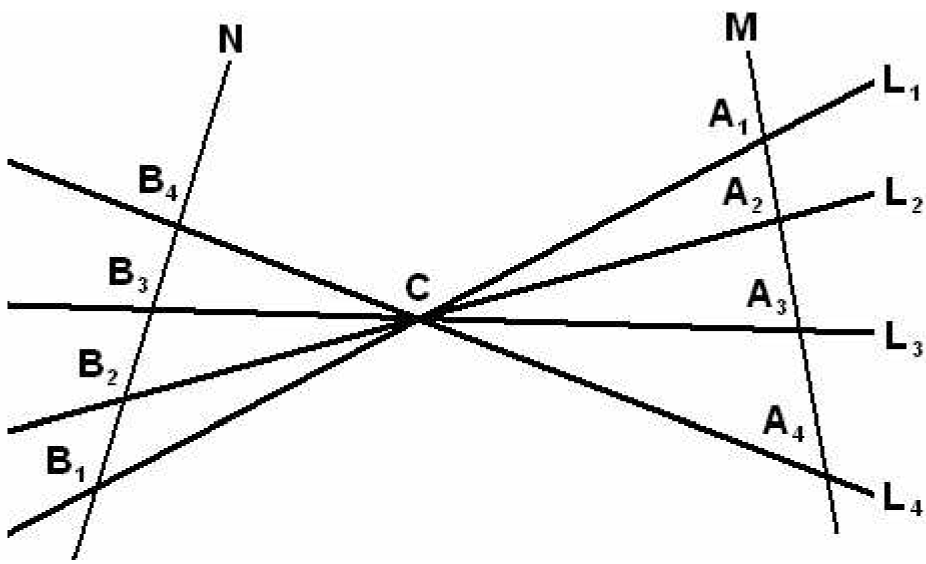


Figure V.1

Proof. Two applications of Theorem 14 imply that

$$\mathsf{XR}(A_1, A_2, A_3, A_4) = \mathsf{XR}(L_1, L_2, L_3, L_4) = \mathsf{XR}(B_1, B_2, B_3, B_4)\ .\blacksquare$$

As noted before, Theorem 15 has a visual application to the interpretation of photographs. Namely, in any photograph of a figure containing four collinear points, the cross ratio of the points is equal to the cross ratio of their photographic images (as before, think of the common point as the aperture of the camera, the line N as the film surface, and the points A_i as the points being photographed).

Finally, we explain the origin of the term *cross ratio*. If V is a vector space over $\mathbb{F}$ and $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are distinct collinear points of V, then the *ratio in which* $\mathbf{c}$ *divides* $\mathbf{a}$ *and* $\mathbf{b}$ is given by $(1-t)/t$, where $\mathbf{c} = t\,\mathbf{a} + (1-t)\,\mathbf{b}$. If $\mathbf{d}$ is a fourth point on the line and $\mathrm{J} : \mathbb{F}^n \to \mathbb{FP}^n$ is the usual inclusion, then Theorem 16 shows that $\mathsf{XR}\big(\mathrm{J}(\mathbf{a}), \mathrm{J}(\mathbf{b}), \mathrm{J}(\mathbf{c}), \mathrm{J}(\mathbf{d})\big)$ is the ratio in which $\mathbf{c}$ divides $\mathbf{a}$ and $\mathbf{b}$ divided by the ratio in which $\mathbf{d}$ divides $\mathbf{a}$ and $\mathbf{b}$.

THEOREM V.16. *Let* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, *and* $\mathbf{d}$ *be distinct collinear points of* $\mathbb{F}^n$, *and write* $\mathbf{c} = t\,\mathbf{a} + (1-t)\,\mathbf{b}$ *and* $\mathbf{d} = s\,\mathbf{a} + (1-s)\,\mathbf{b}$. *Then*

$$\mathsf{XR}\big(\mathrm{J}(\mathbf{a}), \mathrm{J}(\mathbf{b}), \mathrm{J}(\mathbf{c}), \mathrm{J}(\mathbf{d})\big) = \frac{(1-t)\,s}{(1-s)\,t}\ .$$

Proof. We shall identify $\mathbb{F}^{n+1,1}$ with $\mathbb{F}^{n+1}$ and $\mathbb{F}^n \times \mathbb{F}$ in the obvious manner. Recall that $\mathbf{x} \in \mathbb{F}^n$ implies that $\xi = (\mathbf{x}, 1)$ is a set of homogeneous coordinates for $\mathbf{x}$. Clearly we have

$$(\mathbf{c}, 1) \ = \ t\,(\mathbf{a}, 1) \ + \ (1-t)\,(\mathbf{b}, 1)$$

so that we also have

$$(\mathbf{d}, 1) \ = \ \frac{s}{t}\,(t\mathbf{a}, t) \ + \ \frac{1-s}{1-t}\big(\,(1-t)\mathbf{b}, (1-t)\,\big)$$

and the cross ratio formula is an immediate consequence of this.■

The following formula is also useful.

THEOREM V.17. *Let* $\mathbf{a}$, $\mathbf{b}$, *and* $\mathbf{c}$ *and be distinct collinear points of* $\mathbb{F}^n$*, and let* P_∞ *be the point at infinity on the projective extension of this line. If* $\mathbf{c} \ = \ t\mathbf{a} + (1-t)\mathbf{b}$*, then we have*

$$\mathsf{XR}\big(\,\mathrm{J}(\mathbf{a}), \mathrm{J}(\mathbf{b}), \mathrm{J}(\mathbf{c}), P_\infty\,\big) \ = \ \frac{t-1}{t} \ .$$

Proof. Let homogeneous coordinates for J(), $\mathrm{J}(\mathbf{b})$, and $\mathrm{J}(\mathbf{c})$ be given as usual by $(\mathbf{a}, 1)$, $(\mathbf{b}, 1)$, and $(\mathbf{c}, 1)$ respectively. Then $(\mathbf{c}, 1)$ has homogeneous coordinates

$$(\mathbf{c}, 1) \ = \ (t\mathbf{a}, t) \ + \ \big(\,(1-t)\mathbf{b},\ (1-t)\,\big)$$

and P_∞ has homogeneous coordinates

$$(\mathbf{b} - \mathbf{a}, 0) \ = \ (\mathbf{b}, 1) \ - \ (\mathbf{a}, 1) \ = \ \frac{1}{1-t} \cdot \big(\,(1-t)\mathbf{b},\ (1-t)\,\big) \ - \ \frac{1}{t}(t\mathbf{a}, t) \ .$$

Therefore the cross ratio is equal to

$$\frac{-1/t}{1/(1-t)} \ = \ \frac{t-1}{t}$$

which is the formula stated in the theorem.■

EXERCISES

1. Prove Theorem 11.

2. Prove Theorem 12.

3. In the notation of Theorem 13, assume that all the hypotheses except $A_1 \neq B_1$ are valid. Prove that the cross ratio is given by $(x_2 - x_4)/(x_2 - x_3)$ if $A_1 = B_1$.

4. Find the cross ratio of the four collinear points in $\mathbb{RP}^3$ whose homogeneous coordinates are given as follows:

$$\begin{pmatrix} 1 \\ 3 \\ 0 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 7 \\ 4 \\ 4 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 1 \\ -2 \\ -2 \end{pmatrix}$$

5. In $\mathbb{RP}^2$, find the cross ratio of the lines joining the point $\mathrm{J}(\frac{1}{4}, \frac{1}{2})$ to the points with the following homogeneous coordinates:

$$\begin{pmatrix}0\\1\\1\end{pmatrix} \qquad \begin{pmatrix}1\\0\\1\end{pmatrix} \qquad \begin{pmatrix}1\\1\\0\end{pmatrix} \qquad \begin{pmatrix}1\\1\\1\end{pmatrix}$$

6. In $\mathbb{RP}^2$, find the cross ratio formed by the points with homogeneous coordinates ${}^{\mathbf{T}}(1 \quad 1 \quad 2)$ and ${}^{\mathbf{T}}(2 \quad 3 \quad 4)$ and the points in which their line meets the lines defined by $x_1 + x_2 + x_3 = 0$ and $2x_1 + x_3 = 0$.

7. Let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ be distinct collinear points of $\mathbb{R}^n$. Prove that their cross ratio is given by the following formula in which "$\cdot$" denotes the usual vector dot (or inner) product:

$$\mathsf{XR}\big(\,\mathrm{J}(\mathbf{a}), \mathrm{J}(\mathbf{b}), \mathrm{J}(\mathbf{c}), \mathrm{J}(\mathbf{d})\,\big) \;=\; \frac{[\,(\mathbf{b}-\mathbf{c})\cdot(\mathbf{b}-\mathbf{a})\,]\,[\,(\mathbf{d}-\mathbf{a})\cdot(\mathbf{b}-\mathbf{a})\,]}{[\,(\mathbf{c}-\mathbf{a})\cdot(\mathbf{b}-\mathbf{a})\,]\,[\,(\mathbf{d}-\mathbf{d})\cdot(\mathbf{b}-\mathbf{a})\,]}$$

Using this formula, show that the absolute value of the cross ratio is given by the following expression:

$$\frac{|\mathbf{c}-\mathbf{b}|\,|\mathbf{d}-\mathbf{a}|}{|\mathbf{c}-\mathbf{a}|\,|\mathbf{d}-\mathbf{b}|}$$

8. If A, B, C and D are four distinct collinear points of $\mathbb{FP}^n$ (where $\mathbb{F}$ is a commutative field) and $(A',\, B'\, C'\, D')$ is a rearrangement of $(A,\, B,\, C,\, D)$, then by Theorem 12 there are at most six possible values for $\mathsf{XR}(A',\, B',\, C',\, D')$ as $(A',\, B',\, C',\, D')$ runs through all rearrangements. Usually there are exactly six different values for all the possible rearragements, and this exercise analyzes the exceptional cases when there are fewer than six possibilities. By interchanging the roles of $(A,\, B,\, C,\, D)$ if necessary, we can assume that $\mathsf{XR}(A, B, C, D) = r$ is equal to one of the other five expressions in Theorem 12.

(i) Suppose that $1 + 1 \neq 0$ in $\mathbb{F}$ and $\mathsf{XR}(A, B, C, D) = r$ is equal to one of the expressions $1/r$, $1 - r$ or $r/(r-1)$. Prove that r belongs to the set $\{-1, 2, \frac{1}{2}\} \subset \mathbb{F}$ and that the values of the cross ratios $\mathsf{XR}(A',\, B',\, C',\, D')$ for the various rearrangements are precisely the elements of $\{-1, 2, \frac{1}{2}\}$. — Explain why there are three elements in this set if $1 + 1 + 1 \neq 0$ in $\mathbb{F}$ but only one if $1 + 1 + 1 = 0$ in $\mathbb{F}$.

(ii) Suppose that $\mathbb{F}$ is the complex numbers $\mathbb{C}$ and that r is equal to either $1/(1-r)$ or $(r-1)/r$. Prove that r belongs to the set $\{\,\frac{1}{2}1 + \mathbf{i}\sqrt{3},\, \frac{1}{2}(1 - \mathbf{i}\sqrt{3},\,\} \subset \mathbb{C}$ and that the values of the cross ratios $\mathsf{XR}(A',\, B',\, C',\, D')$ for the various rearrangements are precisely the elements of $\{\,\frac{1}{2}1 + \mathbf{i}\sqrt{3},\, \frac{1}{2}(1 - \mathbf{i}\sqrt{3},\,\}$. — Explain why $r^6 = 1$ but $r^m \neq 1$ for $1 \leq m \leq 5$. [*Hint:* Show that $r^2 - r + 1 = 0$ implies $r^3 = -1$ by multiplying both sides by $(r+1)$, and then use this to explain why $r^6 = 1$. In particular, it follows that the possibilities in this case arise if and only if there is some element $r \in \mathbb{F}$ such that $r^6 = 1$ but no smaller positive integral power of r is equal to 1. Such elements exist in $\mathbb{Z}_p$ if p is a prime of the form $6k+1$ — for example, if $p = 7, 13, 19, 31, 37, 43, 61, 67, 73, 79,$ or 97.]

Definitions. In Case (i) of the preceding exercise, there is a rearrangement $(A',\, B'\, C'\, D')$ of $(A,\, B,\, C,\, D)$ such that $\mathsf{XR}(A',\, B',\, C',\, D') = -1$. Ordered quadruples of collinear points satisfying this condition are called *harmonic quadruples*, and they are discussed further in Section V.4, Exercise VI.3.8 and Exercise VII.2.3. In Case (ii), the quadruple is said to form an *equianharmonic set* (the next to last word should be decomposed as equi/an/harmonic). The latter are related to topics in the theory of functions of a complex variable which go far beyond the scope of these notes, and we shall not attempt to give any contexts in which such sets arise.

9. Let $\mathbb{F}$ be a field in which $1+1 \neq 0$ and $1+1+1 \neq 0$, and let A, B, C and D be four distinct collinear points in $\mathbb{FP}^n$.

(*i*) Suppose that $\mathsf{XR}(A,B,C,D) = -1$. Determine all the rearrangements $(A',\ B',\ C',\ D')$ of $(A,\ B,\ C,\ D)$, for which $\mathsf{XR}(A',\ B',\ C',\ D')$ is equal to -1, 2, and $\frac{1}{2}$ respectively.

(*ii*) Suppose that $\mathsf{XR}(A,B,C,D) = r$, where r satisfies the quadratic equation $x^2 - x + 1 = 0$. Explain why $1 - r$ is a second root of the equation such that $r \neq 1 - r$, and using Theorem 12 determine all the rearrangements $(A',\ B',\ C',\ D')$ of $(A,\ B,\ C,\ D)$, for which $\mathsf{XR}(A',\ B',\ C',\ D')$ is equal to r and $(1-r)$ respectively.

10. Let $n \geq 1$, let A be an invertible $(n+1) \times (n+1)$ matrix over $\mathbb{F}$, and let T_A be the geometric symmetry of $\mathbb{FP}^n \cong \mathcal{S}_1(\mathbb{F}^{n+1,1})$ defined by the equation

$$T_A(\mathbf{x}) \;=\; A \cdot \xi \cdot \mathbb{F}$$

where ξ is a set of homogeneous coordinates for $\mathbf{x}$ and the dot indicates matrix multiplication. Suppose that $\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3,\ \mathbf{x}_4$ are collinear points such that the first three are distinct and $\mathbf{x}_4 \neq \mathbf{x}_1$. Prove that T_A preserves cross ratios; more formally, prove that

$$(\mathbf{x}_1,,\, \mathbf{x}_2,\, \mathbf{x}_3,\, \mathbf{x}_4) \;=\; \big(T_A(\mathbf{x}_1),\, T_A(\mathbf{x}_2),\, T_A(\mathbf{x}_3),\, T_A(\mathbf{x}_4)\big)\ .$$

11. Let $\mathbb{F}$ be a field, and let $g : \mathbb{FP}^1 \to \mathbb{FP}^1$ be the 1–1 correspondence such that $g \circ \mathrm{J}(x) \;=\; x^{-1}$ if $0 \neq x \in \mathbb{F}$, and G interchanges the zero point $\mathrm{J}(0)$ and the point at infinity $\infty(\mathbb{FP}^1)$ with homogeneous coordinates given by the transpose of ${}^{\mathbf{T}}(1 \quad 0)$. Prove that there is an invertible 2×2 matrix A such that $g \;=\; T_A$, where the right hand side is defined as in the preceding exercise. [*Hint:* The result extends to *linear fractional transformations* defined by

$$g \circ \mathrm{J}(x) \;=\; \frac{ax+b}{cx+d} \qquad \text{(where} \qquad c \qquad \text{and} \qquad ad \;-\; bc \;\neq\; 0)$$

if $x \;\neq\; -d/c$, while g interchanges $-d/c$ and $\infty(\mathbb{FP}^1)$.]

3. Theorems of Desargues and Pappus

We begin with a new proof of Theorem IV.5 (Desargues' Theorem) for coordinate planes within the framework of coordinate geometry. Among other things, it illustrates some of the ideas that will appear throughout the rest of this chapter. For most of this section, $\mathbb{F}$ will denote an arbitrary skew-fields, and as in the beginning of Section V.1 we shall distinguish between left and right vector spaces.

THEOREM V.18. *A coordinate projective plane $\mathbb{FP}^2$ is Desarguian.*

Proof. Let $\{A,\ B,\ C\}$ and $\{A',\ B',\ C'\}$ be two triples of noncollinear points, and let X be a point which lies on all three of the lines AA', BB' and CC'. Choose homogeneous coordinates $\alpha, \beta, \gamma, \alpha', \beta', \gamma'$ for A, B, C, A', B', C' and ξ for X such that

$$\alpha' \ = \ \xi \ + \ \alpha \cdot a, \qquad \beta' \ = \ \xi \ + \ \beta \cdot b, \qquad \gamma' \ = \ \xi \ + \ \gamma \cdot c.$$

Since $\beta' - \gamma' \ = \ \beta \cdot b \ - \ \gamma \cdot c$, it follows that the point $D \in BC \cap B'C'$ has homogeneous coordinates $\beta' - \gamma'$. Similarly, the points $E \in AC \cap A'C'$ and $F \in AB \cap A'B'$ have homogeneous coordinates $\alpha' - \gamma'$ and $\alpha' - \beta'$ respectively. The sum of these three homogeneous coordinates is equal to zero, and therefore it follows that the points they represent — which are D, E and F — must be collinear.■

Another fundamental result of projective geometry was first stated and proved by Pappus of Alexandria in a Euclidean context.

Pappus' Theorem.[2] *Let $\{A_1,\ A_2,\ A_3\}$ and $\{B_1,\ B_2,\ B_3\}$ be two coplanar triples of non-collinear points in the real projective plane or 3-space. Assume the two lines and six points are distinct. Then the cross intersection points*

$$\begin{array}{lcl} X & \in & A_2B_3 \ \cap \ A_3B_2 \\ Y & \in & A_1B_3 \ \cap \ A_3B_1 \\ Z & \in & A_1B_2 \ \cap \ A_2B_1 \end{array}$$

are collinear.

[2] This is sometimes called the Pappus Hexagon Theorem to distinguish it from the theorems on centroids of surfaces and solids of revolution that were proven by Pappus and rediscovered independently by P. Guldin (1577–1643). A more detailed discussion of Guldin's life and work is available at the following online site: `http:/www.faculty.fairfield.edu/jmac/sj/scientists/guldin.htm` . For reasons discussed in the final section of Chapter VII, the Theorem of Pappus stated in this section is sometimes called *Pascal's Theorem*, and this is especially true for mathematical articles and books written in French or German.

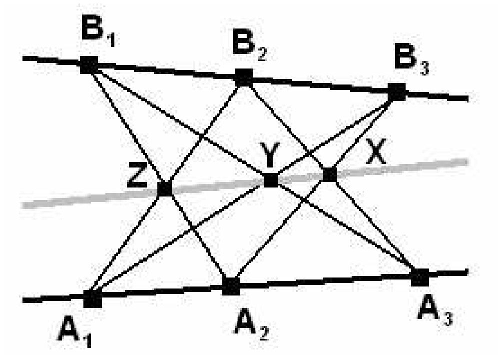


Figure V.2

THEOREM V.19. *Let $\mathbb{F}$ be a skew-field. Then Pappus' Theorem is valid in $\mathbb{FP}^n$ (where $n \geq 2$) if and only if $\mathbb{F}$ is commutative.*

Proof. At most one of the six points $\{A_1,\ A_2,\ A_3,\ B_1,\ B_2,\ B_3\}$ lies on both lines. In particular, there exist distinct numbers $j, k \in \{1,2,3\}$ such that $A_j,\ A_k,\ B_j,\ B_k$ do not lie on both lines. If we re-index the points using a suitable reordering of $\{1,2,3\}$ (explicitly, send j to 1, k to 2, and the remaining number to 3), we find that the renamed points $A_1,\ A_2,\ B_1,\ B_2$ do not lie on both lines. We shall use this revised indexing henceforth. Since the six points are coplanar, by Theorem 6 we may choose homogeneous coordinates α_i and β_j for the points so that $\beta_2 = \alpha_1 + \alpha_2 + \beta_1$. Furthermore, by Theorem 7 we may write $\alpha_3 = \alpha_1 + \alpha_2 \cdot a$ and $\beta_3 = \beta_1 + \beta_2 \cdot b$. Since $Z \in A_1B_2 \cap A_2B_1$, we know there are scalars x, y, u, v such that homogeneous coordinates ζ for Z are given by

$$\zeta \ = \ \alpha_1 \cdot x \ + \ \beta_2 \cdot y \ = \ \beta_1 \cdot u \ + \ \alpha_2 \cdot v \ = \ \alpha_1\,(x+y) \ + \ \alpha_2 \cdot y \ + \ \beta_1 \cdot y \ .$$

Equating coefficients, we see that $x + y = 0$ and $y = u = v$. Hence homogeneous coordinates for Z are given by $\beta_1 + \alpha_2$. Similarly, homogeneous coordinates η for Y are given as follows:

$$\eta \ = \ \alpha_1 \cdot x \ + \ \beta_3 \cdot y \ = \ \alpha_1\,(x + by) \ + \ \beta_1\,(y + by) \ + \ \alpha_2\,(by) \ =$$
$$\alpha_3 \cdot u \ + \ \beta_1 \cdot v \ = \ \alpha_1 \cdot u \ + \ \alpha_2 \cdot au \ + \ \beta_1 \cdot v \ .$$

Equating coefficients as before, we find that homogeneous coordinates for Y are given by

$$\eta \ = \ \alpha_1 \ + \ \alpha_2 \cdot a \ + \ \beta_1\,(1 + b^{-1}a) \ .$$

Still another calculation along the same lines shows that homogeneous coordinates ξ for X are given by

$$\xi \ = \ \alpha_1 \ + \ \alpha_2\,(1 + b^{-1} - ab^{-1}) \ + \ \beta_1\,(1 + b^{-1}) \ .$$

Assume now that $\mathbb{F}$ is commutative. Then $ab^{-1} = b^{-1}a$, and hence

$$\eta \ - \ \xi \ = \ \alpha_2\,(a - 1 - b^{-1} + ab^{-1}) \ + \ \beta_1\,(a + b^{-1}a - 1 - b^{-1})$$

is a scalar multiple of $\alpha_2 + \beta_1$. Since the latter vector is a set of homogeneous coordinates for Z, the conclusion $Z \in XY$ is immediate.$\square$

Conversely, assume that the Pappus' Theorem is always valid in $\mathbb{FP}^n$. It suffices to prove that $ab = ba$ for all $a, b \in \mathbb{F}$ which are not equal to 0 or 1. Let A_1, A_2, B_1, B_2 be four coplanar points, no three of which are collinear, and choose homogeneous coordinates such that $\beta_2 = \alpha_1 + \alpha_2 + \beta_1$. Let $A_3 \in A_1A_2$ and $B_3 \in B_1B_2$ be chosen so that we have homogeneous coordinates of the form $\alpha_3 = \alpha_1 + \alpha_2 \cdot a$ and $\beta_3 = \beta_1 + \beta_2 \cdot b^{-1}$.

Since $Z \in XY$, there exist $x,\ y \in \mathbb{F}$ such that

$$\alpha_2 + \beta_1 = \eta \cdot x + \xi \cdot y .$$

By the calculations in the preceding half of the proof, the right hand side is equal to

$$\alpha_1 (x + y) + \alpha_2 \cdot z + \beta_1 \cdot w$$

where z and w are readily computable elements of $\mathbb{F}$. If we equate coefficients we find that $x + y = 0$ and hence $\alpha_2 + \beta_1 = (\eta - \xi)\, x$. On the other hand, previous calculations show that

$$(\eta - \xi)\, x = \alpha_2 (a - 1 - b - ab)\, x + \beta_1 (a + ba - 1 - b)\, x .$$

By construction, the coefficients of α_2 and β_1 in the above construction are equal to 1. Therefore x is nonzero and

$$a - 1 - b - ab = a + ba - 1 - b$$

from which $ab = ba$ follows.■

Since there exist skew-fields that are not fields (*e.g.*, the quaternions given in Example 3 at the beginning of Appendix A), Theorem 19 yields the following consequence:

> *For each $n \geq 2$ there exist Desarguian projective n-spaces in which Pappus' Theorem is not valid.*

Appendix C contains additional information on such noncommutative skew-fields and their implications for projective geometry. The logical relationship between the statements of Desargues' Theorem and Pappus' Theorem for projective spaces is discussed later in this section will be discussed following the proof of the next result, which shows that Pappus' Theorem is effectively invariant under duality.

THEOREM V.20. *If Pappus' Theorem is true in a projective plane, then the dual of Pappus' Theorem is also true in that plane.*

Proof. Suppose we are given two triples of concurrent lines $\{\alpha_1,\ \alpha_2,\ \alpha_3\}$ and $\{\beta_1,\ \beta_2,\ \beta_3\}$ with distinct points of concurrency that we shall call A and B respectively. Let $C_{i,j}$ denote the common point of α_i and β_j. To prove the planar dual of Pappas' Theorem, we must show that the lines

$$C_{1,3}C_{3,1}, \qquad C_{2,3}C_{3,2}, \qquad C_{1,2}C_{2,1}$$

are concurrent (see the figure below):

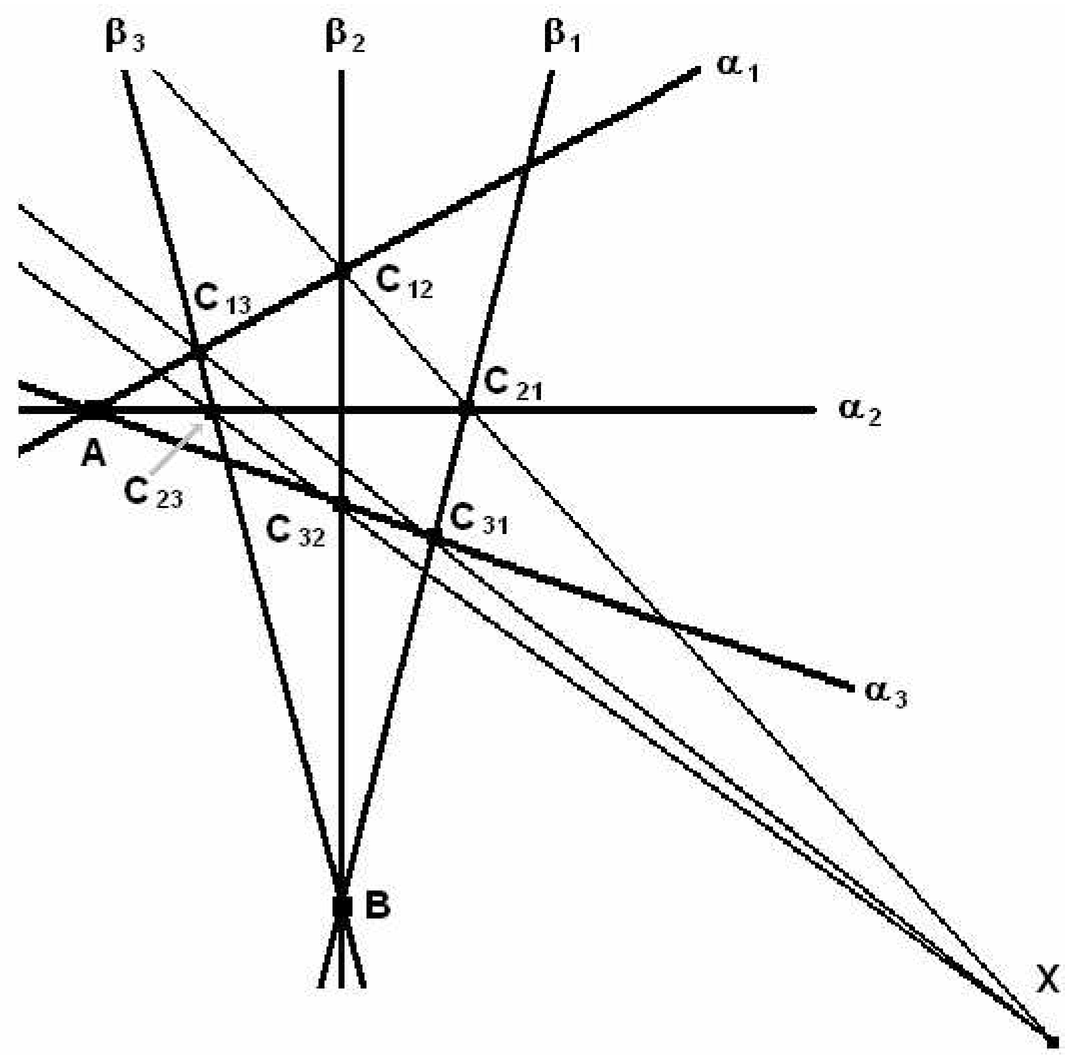


Figure V.3

Now $\{A,\ C_{3,1},\ C_{3,2}\}$ and $\{B,\ C_{2,3},\ C_{1,3}\}$ are two triples of collinear points not on the same line, and hence the three points

$$\begin{aligned} X &\in C_{3,1}C_{2,3} \cap C_{3,2}C_{2,3} \\ C_{2,1} &\in AC_{2,3} \cap BC_{3,1} = \alpha_2 \cap \beta_1 \\ C_{1,2} &\in AC_{1,3} \cap BC_{3,2} = \alpha_1 \cap \beta_2 \end{aligned}$$

are collinear by Pappus' Theorem. Hence $X \in C_{1,2}C_{2,1}$. By the definition of X, it now follows that X lies on all three of the lines $C_{1,3}C_{3,1}$, $C_{2,3}C_{3,2}$, and $C_{1,2}C_{2,1}$.■

A relationship between the validities of Desargues' Theorem and Pappus' Theorem was first formulated by G. Hessenberg (1874–1925). However, his proof was incomplete, and the first correct argument was published by A. Cronheim (1922–2005); the paper containing Cronheim's proof is listed int the bibliography.

THEOREM V.21. *Let $(P, \mathcal{L})$ be a projective plane in which Pappus' Theorem is valid. Then $(P, \mathcal{L})$ is Desarguian.*

The original idea of the proof of Theorem 21 is fairly elementary, but the complete argument is a tedious exercise in the use of such elementary techniques. Details appear on pages 64–66 of the book by Bumcrot listed in the bibliography.

We have already noted that Pappus' Theorem is not necessarily valid in a Desarguian projective plane. However, the following result is true:

THEOREM V.22. *Let P be a* FINITE *Desarguian projective plane. Then Pappus' Theorem is valid in P.*

The main step in the proof of Theorem 22 is an algebraic result of J. H. M. Wedderburn.[3] Proofs appear on pages 375–376 of M. Hall's book on group theory and in the final chapter of the book, *Topics in Algebra*, by I. Herstein (more detailed information appears in the bibliography).

Theorem of Wedderburn. *Every finite skew-field is commutative.*■

Proof of Theorem 22. (*assuming Wedderburn's Theorem*) By Theorem IV.17 we know that P is isomorphic to $\mathbb{FP}^2$ for some skew-field $\mathbb{F}$. Since $\mathbb{FP}^2$ is finite, so is $\mathbb{F}$, and since $\mathbb{F}$ is commutative by Wedderburn's Theorem, Pappus' Theorem is valid by Theorem 19.■

STRENGTHENED RESULT. The theory of finite fields[4] implies that the number of elements in a finite field is p^r, where p is prime and r is a positive integer and also that all finite fields with p^r elements are isomorphic. Combining this with the preceding theorem and a count of the number of elements in $\mathbb{FP}^n$ if $\mathbb{F}$ has $q = p^r$, one can prove the following result:

Complement to Theorem 22. *The number of elements in a finite Desarguian projective n-plane is equal to $1 + q + \cdots + q^n$ where $q = p^r$ for some r, and two finite Desarguian n-spaces with the same numbers of elements are isomorphic.*■

A purely algebraic proof of this result is described in the exercises.

Note on the proofs of Theorem 22 *and its complement.* Since the statements of the result and its complement only involve synthetic and geometric concepts, it is natural to ask if there is a more direct proof that does not require such substantial algebraic input. However, no other proofs are known.

EXERCISES

1. Fill in the omitted details of the calculations for ξ and η in Theorem 19.

[3] JOSEPH H. M. WEDDERBURN (1882–1948) is particularly known for some fundamental results on the structure of certain important types of abstract algebraic systems. As noted in the article by K. H. Parshall in the bibliography, there is a case for attributing the cited result jointly to Wedderburn and L. E. Dickson (1874–1954).

[4] See the books by Herstein or Hungerford.

2. Given three collinear points $\{A,\ B,\ C\}$ and three other collinear points on a different lines, how many different 1–1 ordered correspondences involving the two sets of points are possible?

3. In Exercise 2, assume all points belong to $\mathbb{FP}^2$, where $\mathbb{F}$ is a field. Each correspondence in Exercise 2 determines a line given by Pappus' Theorem. Draw a figure illustrating the lines arising from the two given unordered triples of collinear points. Formulate and prove a statement about these lines.

4. Prove the Complement to Theorem 22. [*Hint:* If $\mathbb{F}$ has q elements, show that the number of nonzero elements of $\mathbb{F}^{n+1}$ is equal to the product of the number of 1-dimensional subspaces times the number of nonzero elements in $\mathbb{F}$. Both of the latter are easy to compute. Solve the resulting equation to obtain the number of elements in $\mathbb{FP}^n$. If two finite fields have different numbers of elements, use the formula to show that their projective n-spaces also have different number of elements because $1 + q + \cdots + q^n$ is a strictly increasing function of q.]

4. Complete quadrilaterals and harmonic sets

In the exercises for the preceding section, we defined the concept of a harmonic quadruple of collinear points. Since the concept plays an important role in this section, we shall repeat the definition and mention a few alternative phrases that are frequently used.

Definition. Let A, B, C, D be a set of four collinear points in $\mathbb{FP}^n$, where $\mathbb{F}$ is a skew-field such that $1+1 \neq 0$ in $\mathbb{F}$. We shall say that the ordered quadruple (A, B, C, D) is a *harmonic quadruple* if $\mathsf{XR}(A,\, B,\, C,\, D) \ = -\ 1$. Frequently we shall also say that *the points A and B separate the points C and D harmonically*,[5] or that the ordered quadruple (A, B, C, D) *forms a harmonic set* (sometimes, with an abuse of language, one also says that the four points form a harmonic set, but Theorem 12 shows that one must be careful about the ordering of the points whenever this wording is used).

Here is the definition of the other basic concept in this section.

Definition. Let A, B, C, D be a set of four coplanar points in a projective incidence space such that no three are collinear. The *complete quadrilateral* determined by these four points, written $\boxtimes ABCD$, is the union of the six lines joining these four points:

$$\boxtimes ABCD \ = \ AB \ \cup \ BC \ \cup \ CD \ \cup \ DA \ \cup \ AC \ \cup \ BD$$

Each line is called a *side*, and the points

$$\begin{array}{l} X \ \in \ AD \ \cap \ BC \\ Y \ \in \ AB \ \cap \ CD \\ Z \ \in \ AC \ \cap \ BD \end{array}$$

are called the *diagonal points* of the complete quadrilateral.

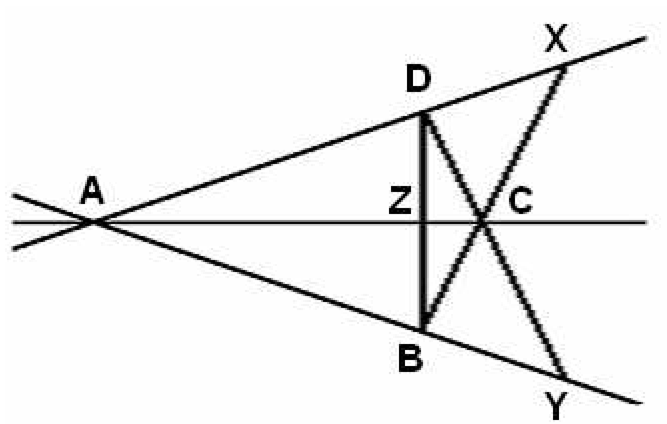


Figure V.4

Note that if the original four points form the vertices of an affine parallelogram in $\mathbb{F}^n$ where $n \geq 2$ and $1+1 \neq 0$ in $\mathbb{F}$, then by Theorem II.26 the first two diagonal points are ideal points but the third is not. On the other hand, if we have a field $\mathbb{F}$ such that $1+1 = 0$ in $\mathbb{F}$, then

[5] This does not depend upon the order of $\{A, B\}$ or $\{C, D\}$ because $\mathsf{XR}(A,\, B,\, C,\, D) \ = -\ 1$ and Theorem 12 imply $\mathsf{XR}(B,\, A,\, C,\, D)$, $\mathsf{XR}(A,\, B,\, D,\, C)$ and $\mathsf{XR}(B,\, A,\, D,\, C)$ are all equal to -1. In fact, the definition is symmetric in the two pairs of points because Theorem 12 implies $\mathsf{XR}(A,\, B,\, C,\, D) \ = \ \mathsf{XR}(C,\, D,\, A,\, B)$.

by Exercise II.4.8 then all three diagonal points are ideal points. The following result, whose significance was observed by G. Fano,[6] is a generalization of these simple observations about parallelograms.

THEOREM V.23. *Let $n \geq 2$, let $\mathbb{F}$ be a skew-field and let $A,\ B,\ C,\ D$ be a set of four coplanar points in $\mathbb{FP}^n$ such that no three are collinear.*

(*i*) *If $1+1 \neq 0$ in $\mathbb{F}$, then the diagonal points of the complete quadrilateral $\boxtimes ABCD$ are* noncollinear.

(*ii*) *If $1+1 = 0$ in $\mathbb{F}$, then the diagonal points of the complete quadrilateral $\boxtimes ABCD$ are* collinear.

Proof. As usual start by choosing homogeneous coordinates $\alpha,\ \beta,\ \gamma,\ \delta$ for $A,\ B,\ C,\ D$ such that $\delta = \alpha+\beta+\gamma$. There exist scalars x, y, u, v such that homogeneous coordinates for X are given by

$$\beta \cdot x \ + \ \gamma \cdot y \ = \ \alpha \cdot u \ + \ \delta \cdot v \ = \ \alpha\,(u+v) \ + \ \beta \cdot v \ + \ \gamma \cdot v \ .$$

Thus we must have $x = y = v$ and $u + v = 0$, so that $\xi = \beta + \gamma$ is a set of homogeneous coordinates for X. Similarly, homogeneous coordinates for Y and Z are given by

$$\eta_0 \ = \ \alpha \cdot x \ + \ \beta \cdot y \ = \ \gamma \cdot u \ + \ \delta \cdot v \ = \ \alpha \cdot v \ + \ \beta \cdot v \ + \ \gamma\,(u+v)$$

and

$$\zeta_0 \ = \ \alpha \cdot x' \ + \ \gamma \cdot y' \ = \ \beta \cdot u' \ + \ \delta \cdot v' \ = \ \alpha \cdot v' \ + \ \beta\,(u'+v') \ + \ \gamma \cdot v'$$

respectively, where $u,\ v,\ u',\ v'$ are appropriate scalars. In these equations we have $x = y = v$ and $x' = y' = v'$, so that Y and Z have homogeneous coordinates $\eta = \alpha+\beta$ and $\zeta = \alpha+\gamma$ respectively. The vectors ξ, η and ζ are linearly independent if $1+1 \neq 0$ in $\mathbb{F}$ and linearly dependent if $1+1 = 0$ in $\mathbb{F}$, and therefore the points X, Y and Z are noncollinear if are linearly independent if $1+1 \neq 0$ in $\mathbb{F}$ and collinear if $1+1 = 0$ in $\mathbb{F}$.■

HYPOTHESIS. Throughout the rest of this section we assume $\mathbb{F}$ is a commutative field; in most but not all cases, we shall also assume that $1 + 1 \neq 0$ in $\mathbb{F}$, but for each result or discussion we shall state explicitly if we making such an assumption.

AFFINE INTERPRETATIONS OF HARMONIC SETS. Given that we have devoted so much attention to affine geometry in these notes, it is natural to ask just what harmonic quadruples look like in affine (and, of course, Euclidean) geometry. Here is one basic result which shows that harmonic quadruples often correspond to familiar concepts in "ordinary" geometry. Additional examples are given in the exercises.

THEOREM V.24. *Let $\mathbb{F}$ be a field in which $1+1 \neq 0$, let $\mathbf{a},\ \mathbf{b},\ \mathbf{c}$be three distinct points in $\mathbb{F}^n$, where $n \geq 1$, and let P_∞ be the ideal point on the projective line $\mathrm{J}(\mathbf{a})\mathrm{J}(\mathbf{b})$. Then $\big(\,\mathrm{J}(\mathbf{a}),\ \mathrm{J}(\mathbf{b}),\ \mathrm{J}(\mathbf{c}),\ \mathrm{J}(P_\infty)\big)$ if and only if $\mathbf{c}$ is the midpoint of $\mathbf{a}$ and $\mathbf{b}$.*

[6] GINO FANO (1871–1952) is recognized for his contributions to the foundations of geometry and to algebraic geometry; an important class of objects in the latter subject is named after him, the projective plane over $\mathbb{Z}_2$ is frequently called the *Fano plane*, and the noncollinearity conclusion in Theorem 23 below is often called *Fano's axiom.*

Proof. According to Theorem 17, if $\mathbf{c} = t\mathbf{a} + (1-t)\mathbf{b}$, then

$$\big(\mathrm{J}(\mathbf{a}),\, \mathrm{J}(\mathbf{b}),\, \mathrm{J}(\mathbf{c}),\, \mathrm{J}(P_\infty)\big) \;=\; -\frac{1-t}{t}\ .$$

It is a straightforward algebraic exercise to verify that the right hand side is equal to -1 if and only if $t = \frac{1}{2}$.■

In view of the preceding theorem, the next result may be interpreted as a generalization of the familiar theorem, "The diagonals of a parallelogram bisect each other" (see Theorem II.26).

THEOREM V.25. *Let $\boxtimes ABCD$ be a complete quadrilateral in $\mathbb{FP}^n$, and let $X,\ Y\ , Z$ be its diagonal points as in the definition. Let $W \in XY \cap BD$ and let $V \in AC \cap XY$. Then we have* $\mathsf{XR}(B, D, W, Z) = \mathsf{XR}(X, Y, W, V) = -1$.

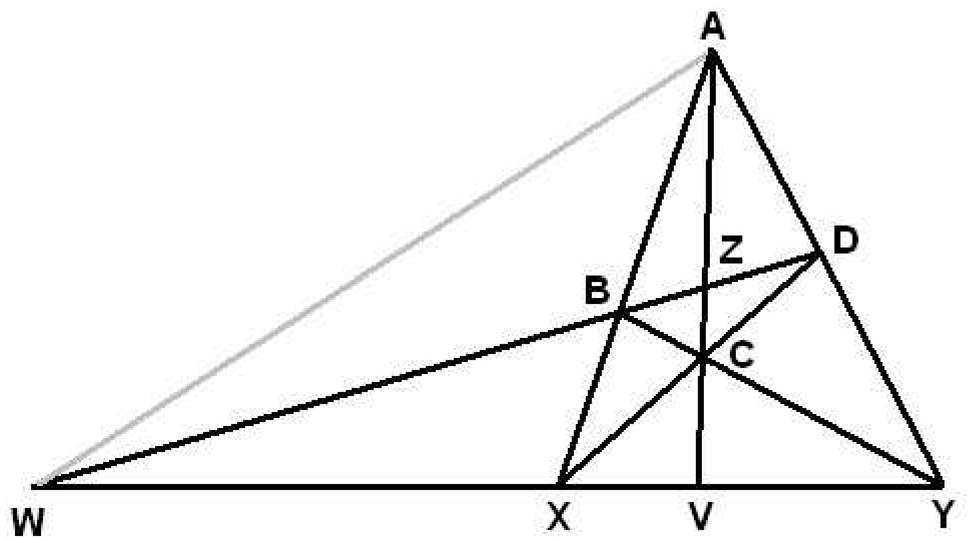


Figure V.5

Proof. Since $X \in AB$, $Y \in AD$, $W \in AW$ and $V \in AZ$, clearly the two cross ratios agree. Choose homogeneous coordinates so that $\delta = \alpha + \beta + \gamma$ (as usual $\alpha,\ \beta,\ \gamma,\ \delta$ are homogeneous coordinates for $A,\ B,\ C,\ D$ respectively). We have already seen that homogeneous coordinates $\xi,\ \eta,\ \zeta$ for $X,\ Y,\ Z$ are given by $\xi = \alpha + \beta$, $\eta = \beta + \gamma$, and $\zeta = \alpha + \gamma$. To find homogeneous coordinates ω for W, note that $\xi + \eta = \alpha + 2\beta + \gamma = \beta + \delta$. Thus $\omega = \alpha + 2\beta + \gamma - \xi + \eta = \beta + \delta$. However, the formulas above imply that

$$\zeta \;=\; \alpha \,+\, \gamma \;=\; -\beta \,+\, \delta$$

and therefore the desired cross ratio formula $\mathsf{XR}(B, D, W, Z) = -1$ follows.■

REMARK. One can use the conclusion of the preceding result to give a purely synthetic definition of harmonic quadruples for arbitrary projective planes. Details appear in many of the references in the bibliography.

Definition. Let L be a line in a projective plane, and let X_i (where $1 \le i \le 6$) be six different points on L. The points X_i are said to form a *quadrangular set* if there is a complete quadrilateral in the plane whose six sides intersect L in the points X_i.

The next result was first shown by Desargues.

THEOREM V.26. *Let $\mathbb{F}$ be a field. Then any five points in a quadrangular set uniquely determine the sixth.*

We should note that the theorem is also true if $\mathbb{F}$ is a skew-field; the commutativity assumption allows us to simplify the algebra in the proof.

Proof. Let L be the given line, and let $\boxtimes ABCD$ be a complete quadrilateral. Define $X_1,\ X_2,\ X_3,\ X_4,\ X_5,\ X_6$ to be the intersections of L with AB, AC, BC, BD, AD, CD respectively. It suffices to prove that X_6 is uniquely determined by the points X_i for $i \leq 5$. The other five cases follow by interchanging the roles of $A,\ B,\ C$ and D.

Choose homogeneous coordinates as usual so that $\delta = \alpha+\beta+\gamma$, and let λ be a set of homogeneous coordinates for L. Define $a = \lambda \cdot \alpha$, $b = \lambda \cdot \beta$, $c = \lambda \cdot \gamma$, and $d = \lambda \cdot \delta$. By construction, we have $d = a + b + c$. Using Theorem 5, we obtain homogeneous coordinates for the points X_i as follows:

$$\begin{aligned}
\xi_1 \ &= \ \lambda \cdot (\alpha \times \beta) \ = \ (\lambda \cdot \beta)\alpha \ - \ (\lambda \cdot \alpha)\beta \ = \ b\alpha \ - \ a\beta \\
\xi_2 \ &= \ c\alpha \ - \ a\gamma \\
\xi_3 \ &= \ c\beta \ - \ b\gamma \\
\xi_4 \ &= \ (a+c)\beta \ - \ b(\alpha+\gamma) \\
\xi_5 \ &= \ (b+c)\alpha \ - \ a(\beta+\gamma) \\
\xi_6 \ &= \ (a+b)\gamma \ - \ c(\alpha+\delta)
\end{aligned}$$

The preceding equations immediately imply that

$$\xi_4 \ = \ \xi_3 \ - \ \xi_1 \ , \qquad \xi_5 \ = \ \xi_1 \ + \ \xi_2 \ , \qquad \xi_6 \ = \ \xi_3 \ - \ \xi_2$$

for the above choices of ξ_i. Furthermore, we have $c\xi_1 - b\xi_2 \ = \ a\xi_3$, and hence we may write the equations above as follows:

$$\xi_3 \ = \ \frac{c}{a}\xi_1 \ - \ \frac{b}{a}\xi_2 \ , \qquad \xi_4 \ = \ \frac{c-a}{a}\xi_1 \ - \ \frac{b}{a}\xi_2 \ , \qquad \xi_6 \ = \ \frac{c}{a}\xi_1 \ - \ \frac{b+a}{a}\xi_2 \ .$$

By definition, the above equations imply the following cross ratio properties:

$$\begin{aligned}
\mathsf{XR}(X_1,\, X_2,\, X_5,\, X_3) \ &= \ -\frac{c}{b} \\
\mathsf{XR}(X_1,\, X_2,\, X_5,\, X_4) \ &= \ -\frac{c-a}{b} \\
\mathsf{XR}(X_1,\, X_2,\, X_5,\, X_6) \ &= \ -\frac{c}{a+b}
\end{aligned}$$

If the first cross ratio is denoted by r and the second by s, then the third is equal to

$$\frac{r}{s-r+1} \ .$$

Thus the third cross ratio only depends upon the points X_i for $i \leq 5$; in particular, if

$$\{X_1,\ X_2,\ X_3,\ X_4,\ X_5,\ Y\}$$

is an arbitrary quadrangular set, then $\mathsf{XR}(X_1,\, X_2,\, X_5,\, Y) \ = \ \mathsf{XR}(X_1,\, X_2,\, X_5,\, X_6)$. By Theorem 10, this implies that $Y = X_6$.$\blacksquare$

Harmonic quadruples and von Staudt's Theorem

The importance of harmonic quadruples in projective geometry is reflected by the following remarkable result of K. von Staudt.[7]

THEOREM V.27. *Let $\varphi : \mathbb{RP}^1 \to \mathbb{RP}^1$ be a $1-1$ onto map which preserves harmonic quadruples; specifically, for all distinct collinear ordered quadruples $(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d})$ we have $\mathsf{XR}(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}) = -1$ if and only if $\mathsf{XR}\left(\varphi(\mathbf{a}),\varphi(\mathbf{b}),\varphi(\mathbf{c}),\varphi(\mathbf{d})\right) = -1$. Then there is an invertible 2×2 matrix A over $\mathbb{R}$ such that φ corresponds to $\mathcal{S}_1(A)$ under the standard identification of $\mathbb{RP}^1$ with $\mathcal{S}_1(\mathbb{R}^{2,1})$, and for all distinct collinear ordered quadruples $(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d})$ we have*

$$\mathsf{XR}(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}) = \mathsf{XR}\left(\varphi(\mathbf{a}),\varphi(\mathbf{b}),\varphi(\mathbf{c}),\varphi(\mathbf{d})\right) .$$

The second part of the conclusion follows from the first (see Exercise V.2.10). One very accessible proof of the first conclusion of the theorem is the following online document:

`http://www-m10.ma.tum.de/~richter/Vorlesungen/ProjectiveGeometrie/Kapitel/Chap5.pdf`

Note that von Staudt's result is only stated for the case $\mathbb{F} = \mathbb{R}$. The corresponding result for more general fields in which $1+1 \neq 0$ is discussed in Section VI.3 following Theorem VI.11 (see the subheading *Collineations of* $\mathbb{FP}^1$).

EXERCISES

1. In $\mathbb{RP}^2$, show that the pair of points whose homogeneous coordinates satisfy

$$x_1^2 - 4x_1x_2 - 3x_2^2 = x_3 = 0$$

separate harmonically the pair whose coordinates satisfy

$$x_1^2 + 2x_1x_2 - x_2^2 = x_3 = 0 .$$

2. Prove that a complete quadrilateral is completely determined by one vertex and its diagonal points.

3. State the plane dual theorem to the result established in Theorem 26.

4. In the Euclidean plane $\mathbb{R}^2$, let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ be noncollinear points, and let the lines bisecting $\angle\,\mathbf{bac}$ and its supplement meet $\mathbf{bc}$ in the points $\mathbf{e}$ and $\mathbf{d}$ of $\mathbb{RP}^2$ respectively. Prove that $\mathsf{XR}(\mathbf{b},\mathbf{c},\mathbf{d},\mathbf{e}) = -1$.

[7] KARL G. C. VON STAUDT (1798–1867) is best known for his contributions to projective geometry and his work on a fundamentally important sequence in number theory called the *Bernoulli numbers*. In his work on projective geometry, von Staudt's viewpoint was highly synthetic, and his best known writings provide a purely synthetic approach to the subject.

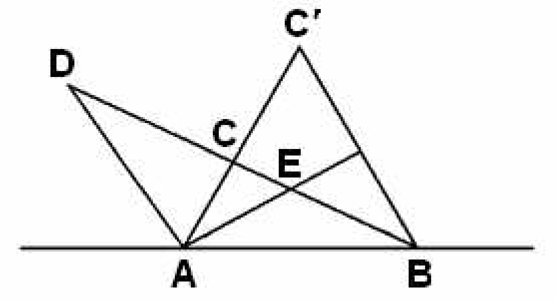


Figure V.6

[*Hint:* Explain why it suffices to consider the triple of points $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}'$ where $\mathbf{c}'$ satisfies $\mathbf{c}' - \mathbf{a} = t(\mathbf{c} - \mathbf{a})$ for some $t > 0$ and $|\mathbf{c}' - \mathbf{a}| = |\mathbf{b} - \mathbf{a}|$. Why is $\mathbf{ae}$ the perpendicular bisector of $\mathbf{b}$ and $\mathbf{c}'$, and why is $\mathbf{bc}'||\mathbf{ad}$? If $\mathbf{e}'$ is the point where $\mathbf{ae}$ meets $\mathbf{bc}'$ and $\mathbf{j}$ is the ideal point on the line $\mathbf{bc}'$, what are $\mathsf{XR}(\mathbf{b}, \mathbf{c}', \mathbf{e}', \mathbf{j})$ and $\mathsf{XR}(\mathbf{b}, \mathbf{c}', \mathbf{j}, \mathbf{e}')$?][8]

5. Let $\mathbb{F}$ be a field in which $1 + 1 \neq 0$, let $A \neq B$ in $\mathbb{FP}^n$, where $n \geq 1$, and let α and β be homogeneous coordinates for A and B. For $1 = 1, 2, 3, 4$ let X_i have homogeneous coordinates $x_i\alpha + \beta$, and assume that $\mathsf{XR}(X_1, X_2, X_3, X_4) = -1$. Prove that

$$\frac{2}{x_4 - x_3} = \frac{1}{x_1 - x_2} + \frac{1}{x_1 - x_4} .$$

6. Let $\mathbb{F}$ be as in the previous exercise, and let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d} \in \mathbb{F}^2$ be distinct points with coordinates $(0, 0)$, $(b, 0)$, $(c, 0)$ and $(d, 0)$ respectively. Assume that $\mathsf{XR}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = -1$. Prove that

$$\frac{1}{b} = \frac{1}{2} \cdot \left(\frac{1}{c} + \frac{1}{d} \right) .$$

[*Hint:* Apply the preceding exercise, taking A and B to be the points U and V whose homogeneous coordinates are ${}^{\mathbf{T}}(1 \quad 0 \quad 0)$ and ${}^{\mathbf{T}}(0 \quad 0 \quad 1)$ respectively.]

> REMARK. The right hand side of the equation above is called the **harmonic mean** of c and d. The harmonic mean was well-known to ancient Greek mathematicians; the name itself[9] was first used by Archytas of Tarentum (*c.* 428 B. C. E. – *c.* 350 B. C. E.), but the concept had been known since the time of the Pythagoreans.

7. Let $\mathbb{F}$ be the real numbers $\mathbb{R}$, let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d} \in \mathbb{R}^2$ be distinct points with coordinates $(0, 0)$, $(b, 0)$, $(c, 0)$ and $(d, 0)$ respectively, and assume that b, c and d are all positive. Prove that $\mathsf{XR}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$ is positive if either b is less than or greater than both c and d, but $\mathsf{XR}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$ is negative if either $0 < c < b < d$ or $0 < d < b < c$. — In other words, the cross ratio is negative if and only if the points 0 and b *separate* the points c and d in the sense that one of the latter lies on the bounded open interval defined by $0 < x < b$ and the other lies on the unbounded interval defined by $x > b$. More will be said about this concept of separation in the final section of Chapter VI. [*Hint:* Use the same methods as in the preceding exercise to express the cross ratio in terms of the coordinates of the four points.]

---

[8] See Moise, Prenowitz and Jordan, or the online site mentioned in the Preface for mathematically sound treatments of the Euclidean geometry needed for this exercise.

[9] More correctly, the corresponding name in Classical Greek.

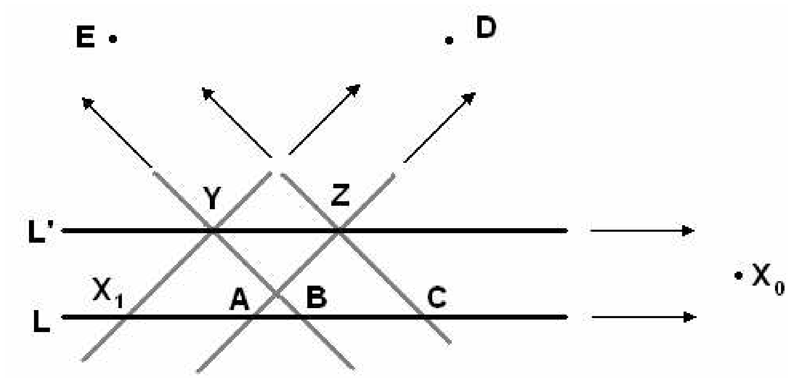
5. Interpretation of addition and multiplication

Theorem IV.15 states that if an n-dimensional projective incidence space is isomorphic to $\mathbb{FP}^n$ for some skew-field $\mathbb{F}$, then the latter is unique up to algebraic isomorphism. In particular, if $\mathbb{E}$ is a skew-field such that $\mathbb{EP}^n$ is isomorphic to $\mathbb{FP}^n$, then $\mathbb{E}$ is algebraically isomorphic to $\mathbb{F}$. The reason for this is that addition and multiplication in the underlying skew-field have synthetic interpretations in terms of certain geometric constructions which are motivated by ordinary Euclidean geometry. If the two coordinate projective spaces as above are isomorphic, this means that the algebraic operations in each are characterized by the same synthetic constructions, and therefore the algebraic operations in the two underlying skew-fields must be isomorphic.

We shall take the preceding discussion further in Chapter VI, where we shall use the synthetic interpretations of addition and multiplication to give an complete description of all geometrical (incidence space) automorphisms of $\mathbb{FP}^n$, where $\mathbb{F}$ is a field and $n \geq 2$.

In order to simplify the algebra, we again restrict attention to commutative fields; however, all the results in this section are equally valid for arbitrary skew-fields (and the result on automorphisms in Chapter VI also extend to the noncommutative case).

Euclidean addition of lengths. If L is a line in the Euclidean plane containing the points X_1, A and B such that the lengths of the segments $[X_1A]$ and $[X_1B]$ have lengths a and b respectively, then the figure below indicates one method for finding a point $C \in L$ such that the segment $[X_1C]$ has length $a + b$.



$L||L'$, $X_1Y||AZ$, $YB||ZC$ with ideal points X_0, E and D respectively.

Figure V.7

This example motivates the following abstract result:

Theorem V.28. *Let $\mathbb{F}$ be a field, and let $n \geq 2$. Let L be a line in $\mathbb{FP}^n$ containing a point X_0, let M be another line containing X_0, let X_1 be another point on L, and let U be a third point on $L = X_0X_1$. Let A and B be points of $L - \{P_0\}$, and let Y be a point in the plane of L and M which does not line on either line. Let D be the (unique) intersection point of $X_1Y \cap M$ (note that $D \neq X_1$, for that would imply X_1 and X_0 both lie on $L \cap M$), and let*

$$Z \in AD \cap X_0Y , \qquad E \in YB \cap M , \qquad C \in RE \cap L .$$

Then $C \neq X_0$ and $\mathsf{XR}(X_0, X_1, U, C) = \mathsf{XR}(X_0, X_1, U, A) + \mathsf{XR}(X_0, X_1, U, B)$.

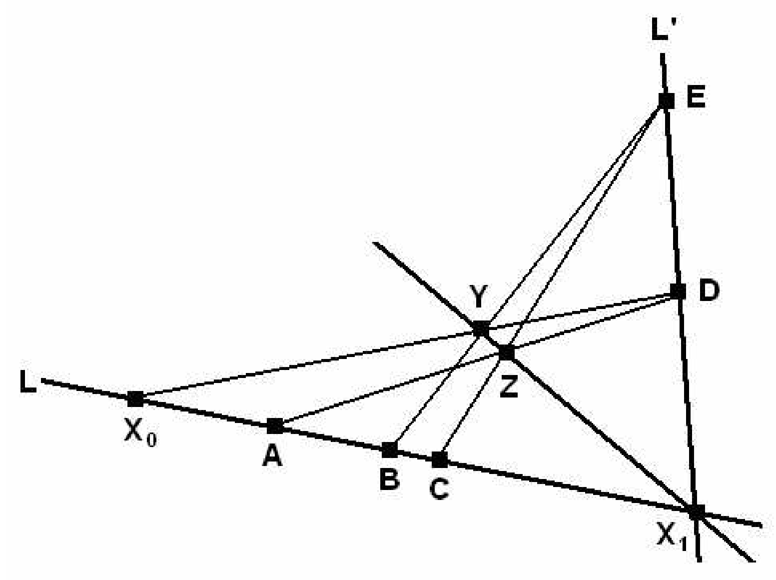


Figure V.8

Proof. Let V be the point where X_0Y meets DU. Then no three of the points $\{X_0,\ X_1,\ D,\ V\}$ are collinear. Choose homogeneous coordinates $\xi_0,\ \xi_1,\ \delta,\ \psi$ for $X_0,\ X_1,\ D,\ V$ so that $\psi\ =\ \xi_0+\xi_2+\delta$. Since $Y\in DX_1\cap X_0V$ it follows that $\eta=\xi_1+\delta$, and since $U\in X_0Y\cap DV$ it follows that $\eta=\xi_1+\delta$.

By the definition of cross ratios, there are homogeneous coordinates α and β for A and B respectively such that $\alpha\ =\ a\xi_0+\xi_1$ and $\beta\ =\ b\xi_0+\xi_1$, where $a=\mathsf{XR}(X_0,X_1,U,A)$ and $b=\mathsf{XR}(X_0,X_1,U,B)$. Since $Z\in AD\cap X_0Y$, there exist $x,\ y,\ u,\ v\in\mathbb{F}$ such that homogeneous coordinates for ζ are given by

$$x\alpha\ +\ y\cdot a\ =\ u\xi_0\ +\ v\eta\ .$$

Using the preceding equations, this equation may be expanded to

$$xa\xi_0\ +\ x\xi_1\ +\ y\delta\ =\ u\xi_0\ +\ v\xi_1\ +\ v\delta\ .$$

Therefore we must have $x=v$ and $xa=u$, so that $\zeta\ =\ a\xi_0+\xi_1+\delta$. Similarly, $D\in YB\cap X_0D$ implies an equation of the form $x\eta+y\beta\ =\ u\xi_0+\delta$, which is equivalent to

$$x\xi_1\ +\ x\delta\ +\ yb\xi_0\ +\ y\xi_1\ =\ u\xi_0\ +\ v\delta\ .$$

Therefore $yb=u$, $x=v$ and $x+y=0$ imply that homogeneous coordinates ε for E are given by

$$\varepsilon\ =\ -b\xi_0\ +\ \delta\ .$$

Finally, $C\in ZE\cap X_0X_1$ implies an equation

$$x\zeta\ +\ y\varepsilon\ =\ u\xi_0\ +\ v\xi_1$$

which is equivalent to

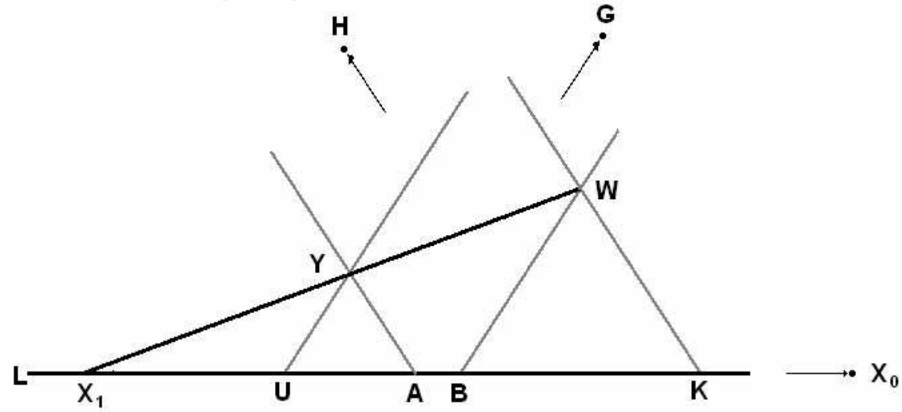
$$xa\xi_0\ +\ x\xi_1\ +\ x\delta\ -\ y\delta\ -\ yb\xi_0\ =\ u\xi_0\ +\ v\xi_1\ .$$

Thus $x+y=0$, $xa-yb=u$ and $x=v$ imply that homogeneous coordinates γ for C are given by

$$\gamma\ =\ (a+b)\xi_0\ +\ \xi_1\ .$$

In particular, it follows that $C\neq X_0$ and $\mathsf{XR}(X_0,X_1,U,C)\ =\ a+b.$■

Euclidean multiplication of lengths. Similarly, if L is a line in the Euclidean plane containing the points X_1, U, A and B such that the lengths of the segments $[X_1U]$, $[X_1A]$, and $[X_1B]$ have lengths 1, a and b respectively, then the figure below indicates one method for finding a point $K \in L$ such that the segment $[X_1K]$ has length $a \cdot b$.



$YA||WK$ and $YU||BW$ with ideal points H and G respectively, and X_0 is the ideal point of L.

Figure V.9

Here is the corresponding abstract result:

THEOREM V.29. *Let L, X_0, X_1, U, A, B, Y, D satisfy the conditions of the previous theorem, and let*

$$G \in YU \cap L , \qquad H \in YA \cap L , \qquad W \in X_1D \cap BG , \qquad K \in HW \cap X_0X_1 .$$

Then $K \neq X_0$ and $\mathsf{XR}(X_0, X_1, U, C) = \mathsf{XR}(X_0, X_1, U, A) \cdot \mathsf{XR}(X_0, X_1, U, B)$.

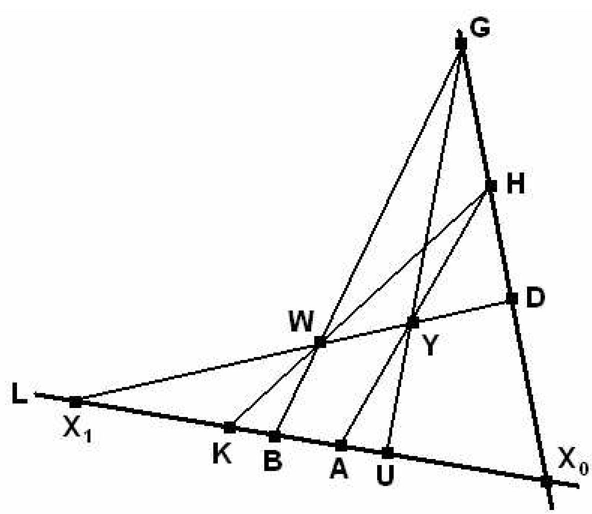


Figure V.10

Proof. The problem here is to find homogeneous coordinates for G, H, W and K. Unless otherwise specified, we shall use the same symbols for homogeneous coordinates representing

$L,\ X_0,\ X_1,\ A,\ B,\ Y,\ D$ as in the previous argument, and choose homogeneous coordinates for U of the form $\xi_0 + \xi_1$. Since $G \in YU \cap X_0D$, there is an equation of the form

$$x\eta \ + \ y(\xi_0 + \xi_1) \ = \ u\xi_0 \ + \ v\delta$$

which is equivalent to

$$x(\xi_1 + \delta) \ + \ y(\xi_0 + \xi_1) \ = \ u\xi_0 \ + \ v\delta \ .$$

Therefore $y = u$, $x + y = 0$, and $x = v$ imply that homogeneous coordinates χ for G are given by

$$\chi \ = \ \xi_0 - \delta \ .$$

Since $H \in YA \cap X_0D$, there is an equation of the form

$$x\xi_0 \ + \ y\alpha \ = \ u\xi_0 \ + \ v\delta$$

which is equivalent to

$$x(\xi_0 + \delta) \ = \ y(a\xi_0 + \xi_1) \ = \ u\xi_0 \ + \ v\delta \ .$$

Therefore $ya = u$, $x + y = 0$, and $x = v$ imply that homogeneous coordinates θ for H are given by

$$\theta \ = \ -a\xi_0 \ + \ \delta \ .$$

Since $W \in X_1U \cap BG$, there is an equation

$$x\xi_1 \ + \ y\delta \ = \ u\beta \ + \ v\chi$$

which is equivalent to

$$x\xi_1 \ + \ y\delta \ = \ u(b\xi_0 + \xi_1) \ + \ v(\xi_0 - \delta) \ = \ (ub + v)\xi_0 \ + \ u\xi_1 \ - \ v\delta \ .$$

Thus $ub + v = 0$, $-v = y$, and $u = x$ imply that homogeneous coordinates ω for W are given by

$$\omega \ = \ \xi_1 \ - \ \delta \ .$$

Finally, $K \in HW \cap X_0X_1$ implies an equation of the form

$$x\theta \ + \ y\omega \ = \ u\xi_0 \ + \ v\xi_1$$

which is equivalent to

$$u\xi_0 \ + \ v\xi_1 \ = \ x(-a\xi_0 + \delta) \ + \ y(\xi - b\delta) \ = \ -ax\xi_0 \ + \ y\xi_1 \ - \ (ax + by)\delta \ .$$

Therefore $u = -ax$, $v = y$.and $x + by = 0$ imply

$$u \ = \ -ax \ = \ -a(-by) \ = \ aby \ = \ abv$$

and hence homogeneous coordinates κ for K are given by $\kappa \ = \ ab\xi_0 \ + \ \xi_1$.■

CHAPTER VI

MULTIDIMENSIONAL PROJECTIVE GEOMETRY

In this chapter we shall study coordinate projective spaces of arbitrary dimension. As in the previous chapter, we shall use concepts from linear algebra extensively. Although some portions of this chapter contain results of the previous one as special cases, most of the material involves concepts not covered earlier in these notes.

One major difference between this chapter and the previous one is that we are mainly interested in somewhat different types of results. In particular, we are interested in the geometric automorphisms of a coordinate projective space $\mathbb{FP}^n$, and the results of this chapter give a simple but complete description of them. In the final section of this chapter we shall assume that we have a field (or skew-field) of scalars $\mathbb{F}$ which has a notion of ordering with the same basic properties of the orderings of the real or rational numbers, and we shall analyze the geometrical implications of such algebraic orderings.

1. Linear varieties and bundles

Our first objective is to extend the results of Section V.1 on duality and homogeneous coordinates from $\mathbb{FP}^2$ to $\mathbb{FP}^n$, where $n \geq 3$ is arbitrary. As indicated in Theorem IV.16, if (S, Π, d) is an n-dimensional projective (incidence) space, then the "points" of the dual projective n-space (S^*, Π^*, d^*) are the hyperplanes of S. Suppose now that $S = \mathbb{FP}^n$ for some skew-field $\mathbb{F}$; by the results of Section V.1, if $n = 2$ then we can introduce homogeneous coordinates into the dual projective plane $\left(\mathbb{FP}^2\right)^*$. We shall extend this to all $n \geq 2$, showing that one can define well-behaved homogeneous coordinates for the hyperplanes of $\mathbb{FP}^n$ for all $n \geq 2$ such that most of the fundamental results from Section V.1 also extend to this more general setting.

According to Theorem III.12, a hyperplane in $\mathbb{FP}^n$ is definable by a right homogeneous linear equation

$$\sum_{i=1}^{n+1} u_i \, x_i \ = \ 0$$

where the coefficients u_i are not all zero. Furthermore, two n-tuples $(u_1, \ \cdots \ , u_{n+1})$ and $(v_1, \ \cdots \ , v_{n+1})$ define the same hyperplane if and only if there is a nonzero $k \in \mathbb{F}$ such that $u_i \ = \ k \, v_i$ for all i (compare Section V.1). This immediately yields the following analog of Theorem V.1:

THEOREM VI.1. *Let $\mathbb{F}$ be a skew-field, and let $n \geq 2$. Then the set of hyperplanes in $\mathbb{FP}^n$ is in $1-1$ correspondence with $\mathcal{S}_1(\mathbb{F}^{1,n+1})$. Furthermore, if the hyperplane H corresponds to the left 1-dimensional vector subspace $\mathbb{F} \cdot \theta$ and $X \in \mathbb{FP}^n$ is given by $\xi \cdot \mathbb{F}$, then $X \in H$ if and only if $\theta \cdot \xi \ = \ 0$.*■

As before, if the hyperplane H corresponds to the left 1-dimensional vector subspace Ω of $\mathbb{F}^{1,n+1}$, then a set of *homogeneous coordinates* for H is any nonzero vector in Ω.

Motivated by the preceding description of hyperplanes, we define a *linear variety* in $\mathbb{FP}^n$ to be the set of all points whose homogeneous coordinates satisfy a system of linear homogeneous equations

$$\sum_{j=1}^{n+1} u_{i.j} x_j \ = \ 0 \qquad\qquad 1 \ \leq \ i \ \leq \ m \ .$$

The following result shows that linear varieties are the same as geometrical subspaces.

THEOREM VI.2. *Let V be a linear variety defined by a system of linear homogeneous equations as above, and suppose that the (left) row rank of the matrix $B \ = \ (u_{i,j})$ is equal to r. Then V is an r-plane in $\mathbb{FP}^n$.*

Proof. If V_0 is the solution space of the system of equations, then clearly $V \ = \ \mathcal{S}_1(V_0)$. Since the rank of B is r, then dimension of V_0 is equal to $n+1-r$ by Theorem A.10, and hence V is an $(n-r)$-plane in $\mathbb{FP}^n$.

On the other hand, assume that W is a $(k+1)$-dimensional vector subspace of $\mathbb{F}^{n+1,1}$, so that $\mathcal{S}_1(W)$ is a k-plane. Let $\mathbf{w}_1, \ \cdots \ \mathbf{w}_{k+1}$ be a basis for W, and write $\mathbf{w}_i \ = \ {}^{\mathbf{T}}(w_{1,i} \quad \cdots \quad w_{k+1,i})$. Consider the left-homogeneous system of linear equations

$$\sum_i y_i \, w_{j,i} \ = \ 0 \qquad\quad (1 \ \leq \ i \ \leq \ k+1) \ .$$

Since the right column rank of the matrix $C \ = \ (w_{j,i})$ is equal to k+1, the left subspace of solutions has dimension equal to $n-k$ (again by Theorem A.10). Let $\mathbf{v}_1, \ \cdots \ \mathbf{v}_{n-k}$ be a basis for the space of solutions, and write $\mathbf{v}_i \ = \ (v_{i,1} \quad \cdots \quad v_{i,n+1})$. Then, by construction, the vector subspace W is contained in the space of solutions of the system

$$\sum_j v_{j,i} \, x_j \ = \ 0 \qquad\quad (1 \ \leq \ j \ \leq \ n-k) \ .$$

On the other hand, the space of solutions W' has dimension equal to

$$(n+1) \ - \ (n-k) \ = \ k \ + \ 1 \ .$$

Since this is the dimension of W, we must have $W' = W$, and this proves the second half of the theorem.■

Similarly, we may define a *linear variety of hyperplanes* to be the set of all hyperplanes whose homogeneous coordinates satisfy a system of left-homogeneous linear equations

$$\sum_i u_i \, x_{i,j} \ = \ 0 \qquad\quad (1 \ \leq \ j \ \leq \ m) \ .$$

If the right rank of $X \ = \ (x_{i,j})$ is r, the variety of hyperplanes is said to be $(n-r)$-dimensional. The following result shows that linear varieties of hyperplanes are also equivalent to geometrical subspaces of $\mathbb{FP}^n$.

THEOREM VI.3. *An r-dimensional linear variety of hyperplanes in $\mathbb{FP}^n$ consists of all hyperplanes containing a fixed $(n-r-1)$-plane in the terminology of Chapter* IV*, a linear bundle with the given $(n-r-1)$-plane as center). Conversely, every $(n-r-1)$-plane in $\mathbb{FP}^n$ is the center of some linear variety of hyperplanes.*

Proof. The ideas are similar to those employed in Theorem 2. Let C_0 be the span of the rows of the matrix $(x_{i,j})$. By hypothesis, $\dim C_0 = n-r$. Thus $C = \mathcal{S}_1(C_0)$ is an $(n-r-1)$-plane in $\mathbb{FP}^n$, and every hyperplane containing it automatically belongs to the linear variety. Conversely, if $\mathbf{y}_0 \in C_0$, then we may write $\mathbf{y} \;=\; \sum_i \mathbf{x}_i \, r_i$ where $\mathbf{x}_i \;=\; (x_{1,i}, \; \cdots \; , x_{n+1,i})$ and $r_i \in \mathbb{F}$, so that if H lies in the variety and θ is a set of homogeneous coordinates for H then we have

$$\theta \cdot \mathbf{y} \;=\; \sum_i \, (\theta \cdot \mathbf{x}_i) \, r_i \;=\; 0 \; .$$

Thus every hyperplane in the variety contains every point of C. This proves the first half of the theorem.

Now suppose that we are given an $(n-r-1)$-plane $Z \;=\; \mathcal{S}_1(Z_0)$. Let $\mathbf{z}_1, \; \cdots \; , \mathbf{z}_{n-r}$ be a basis for Z_0, and write $\mathbf{z}_j \;=\; (z_{1,j}, \; \cdots \; , z_{n+1,j})$. Consider the variety of hyperplanes defined by the system of homogeneous equations

$$\sum_j \; u_i \, z_{i,j} \;=\; 0 \qquad\qquad (1 \;\leq\; j \;\leq\; n-r) \; .$$

Since the right rank of the matrix $(z_{i,j})$ is equal to $(n-r)$, this bundle is r-dimensional. Furthermore, its center Z' is an $(n-r-1)$-plane which contains every point of Z by the reasoning of the previous paragraph. Therefore we have $Z \;=\; Z'$.■

The preceding result has some useful consequences.

Theorem VI.4. *Let $(\mathbb{FP}^n)^*$ be the set of hyperplanes in $\mathbb{FP}^n$, and let Π^* and d^* be defined as in Section* IV.3. *Then a subset $S \subset (\mathbb{FP}^n)^*$ is in Π^* if and only if it is a linear variety of hyperplanes, in which case $d^*(S)$ is the dimension as defined above.*■

Theorem VI.5. (compare Theorem V.1) *The triple*

$$\big((\mathbb{FP}^n)^* , \; \Pi^*, \; d^* \big)$$

is a projective n-space which is isomorphic to $\mathcal{S}_1(\mathbb{F}^{1,n+1})$.

Since Theorem 4 is basically a restatement of Theorem 3, we shall not give a proof. However, a few remarks on Theorem 5 are in order.

Proof of Theorem 5. By Theorem 1 we have a 1–1 correspondence between $(\mathbb{FP}^n)^*$ and $\mathcal{S}_1(\mathbb{F}^{1,n+1})$. Furthermore, the argument used to prove Theorem 2 shows that r-dimensional varieties of hyperplanes correspond to set of the form $\mathcal{S}_1(V)$, where V is an $(r+1)$-dimensional left subspace of $\mathbb{F}^{1,n+1}$ (merely interchange the roles of left and right in the proof, switch the orders of the factors in products, and switch the orders of double subscripts). But r-dimensional linear bundles correspond to r-dimensional linear varieties of hyperplanes by Theorems 3 and 4. Combining these, we see that r-dimensional linear bundles of hyperplanes correspond to r-planes in $\mathcal{S}_1(\mathbb{F}^{1,n+1})$ under the 1–1 correspondence between $(\mathbb{FP}^n)^*$ and $\mathcal{S}_1(\mathbb{F}^{1,n+1})$.■

By the Coordinatization Theorem (Theorem IV.18), this result implies the first half of Theorem IV.16. On the other hand, if we interchange the roles of left and right, column vectors and row vectors, and the orders of multiplication and indices in the reasoning of this section, we find that the dual of $\mathcal{S}_1(\mathbb{F}^{1,n+1})$ is isomorphic to $\mathbb{FP}^{n+1}$. In fact, this isomorphism $h : \mathcal{S}_1(\mathbb{F}^{1,n+1})^* \to \mathcal{S}_1(\mathbb{F}^{n+1,1})$ is readily seen to have the property that the composite $h \circ f^* \circ e$ given by

$$\mathbb{FP}^n \xrightarrow[\cong]{e} (\mathbb{FP}^n)^{**} \xrightarrow{f^*} \mathcal{S}_1(F^{1,n+1})^* \xrightarrow{h} \mathbb{FP}^n$$

is the identity (here f^* is an isomorphism of dual spaces induced by f as in Exercise IV.3.4). This establishes the second half of Theorem IV.16 and allows us to state the **Principle of Duality in Higher Dimensions:**

METATHEOREM VI.6. *A theorem in projective geometry in dimension $n \geq 2$ remains true if we interchange the expressions* point *and* hyperplane*, the phrases* r-planes in an n-space *and* $(n-r-1)$-planes in an n-space*, and the words* contains *and* is contained in.■

We shall now assume that $\mathbb{F}$ is commutative. Since $\mathbb{F} = \mathbb{F}^{\mathsf{OP}}$ in this case, the dual of $\mathbb{FP}^n$ is isomorphic to $\mathbb{FP}^n$. Hence the metatheorem may be modified in an obvious way to treat statements about projective n-spaces **over fields**.

The cross ratio of four hyperplanes four hyperplanes in $\mathbb{FP}^n$ containing a common $(n-2)$-plane may be defined in complete analogy with the case $n = 2$, which was treated in Section V.2. In particular, Theorem V.14 generalizes as follows.

THEOREM VI.7. *Let H_1, H_2, H_3 be distinct hyperplanes through an $(n-2)$-plane K in $\mathbb{FP}^n$, and let $H_4 \neq H_1$ be another hyperplane through K. Let L be a line disjoint from K, and let X_i be the unique point where L meets H_i for $1 = 1, 2, 3$. Then the point $X_4 \in L$ lies on H_4 if and only if we have*

$$\mathsf{XR}(X_1, X_2, X_3, X_4) \;=\; \mathsf{XR}(H_1, H_2, H_3, H_4)\ .$$

The proof of this result is formally identical to the proof of Theorem V.1.■

EXERCISES

1. Let $\mathbb{F}$ be a field, and let $X,\ Y,\ Z \in \mathbb{FP}^3$ be noncollinear points. Suppose that homogeneous coordinates for these points are respectively given as follows:

$$\xi \;=\; \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \qquad \eta \;=\; \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \qquad \zeta \;=\; \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}$$

Prove that the plane they determine has the following homogeneous coordinates:

$$\left(\begin{vmatrix} x_2 & x_3 & x_4 \\ y_2 & y_3 & y_4 \\ z_2 & z_3 & z_{,4} \end{vmatrix},\ - \begin{vmatrix} x_1 & x_3 & x_4 \\ y_1 & y_3 & y_4 \\ z_1 & z_3 & z_{,4} \end{vmatrix},\ \begin{vmatrix} x_1 & x_2 & x_4 \\ y_1 & y_2 & y_4 \\ z_1 & z_2 & z_4 \end{vmatrix},\ - \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix} \right)$$

By Theorem A.11, not all of the four determinants vanish because ξ, η and ζ are linearly independent. To see that $X,\ ,Y,\ Z$ lie on the above hyperplane, consider the determinants of the three 4×4 matrices whose rows are given by ${}^{\mathbf{T}}\omega$, ${}^{\mathbf{T}}\xi$, ${}^{\mathbf{T}}\eta$ and ${}^{\mathbf{T}}\zeta$, where ω runs through the three vectors in the set $\xi,\ \eta,\ \zeta$.

2. Explain why four points $\mathbf{p}_i = (x_i, y_i, z_i) \in \mathbb{F}^3$ (where $1 \le i \le 4$) are coplanar if and only if the 4×4 determinant

$$\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix}$$

is zero, where $\mathbb{F}$ is a field. Formulate an analogous statement for n dimensions. [*Hint:* For both parts, use the properties of determinants as described in Appendix A and the characterization of hyperplanes in terms of n-dimensional vector subspaces of $\mathbb{F}^{1,n+1}$.]

3. Write out the 3-dimensional projective duals of the following concepts:

(a) A set of collinear points.

(b) A set of concurrent lines.

(c) The set of all planes through a given point.

(d) Four coplanar points, no three of which are collinear.

(e) A set of noncoplanar lines.

4. What is the 3-dimensional dual of Pappus' Theorem?

5. Let $\{A,\ B,\ C,\ D\}$ and $\{A',\ B',\ C'',D'\}$ be two triples of noncoplanar points in a projective 3-space, and assume that the lines AA', BB', CC' and DD' are concurrent. Prove that the lines of intersection

$$\begin{aligned} G &= \mathsf{plane}(ABC) \ \cap\ \mathsf{plane}(A'B'C') \\ H &= \mathsf{plane}(ABD) \ \cap\ \mathsf{plane}(A'B'D') \\ K &= \mathsf{plane}(ACD) \ \cap\ \mathsf{plane}(A'C'D') \\ L &= \mathsf{plane}(BCD) \ \cap\ \mathsf{plane}(B'C'D') \end{aligned}$$

are coplanar, and state and prove the converse.

6. Find the equations of the hyperplanes through the following quadruples of points in $\mathbb{RP}^4$.

(a)

$$\begin{pmatrix} 2 \\ 3 \\ 1 \\ 0 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 4 \\ 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \\ 0 \end{pmatrix}$$

(b)

$$\begin{pmatrix} 3 \\ 4 \\ 0 \\ 0 \\ 2 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} 2 \\ 0 \\ 5 \\ 1 \\ 2 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \\ 0 \end{pmatrix}$$

2. Projective coordinate systems

Theorems V.6 and V.7, which provided particularly convenient choices for homogeneous coordinates in one or two dimensions, proved to be extremely useful in Chapter V. We shall prove a corresponding result for higher dimensions in this section; as one might expect, this result has corresponding importance in higher dimensional projective geometry. All the results of this section (except for Exercise 1) are valid if $\mathbb{F}$ is an arbitrary skew-field.

THEOREM VI.8. *Let $\{B_0, \cdots, B_r\}$ be a set of independent points in $\mathbb{FP}^n$, and let U be a point in the r-plane $B_0 \cdots B_r$ such that every proper subset of $\{B_0, \cdots, B_r, U\}$ is independent. Then homogeneous coordinates β_i and ψ can be chosen for these points may be chosen such that*

$$(\ddagger) \qquad \psi = \beta_0 + \cdots + \beta_r .$$

Furthermore, if β_i' and ψ' is another collection of homogeneous for these points such that $(\ddagger)$ holds, then there is a nonzero $a \in \mathbb{F}$ such that $\psi = \psi' a$ and $\beta_i = \beta_i' a$ for $i = 0, \cdots, r$.

Proof. . Since the points B_i are independent, if we take arbitrary homogeneous coordinates $\widetilde{\beta_i}$ and ψ then there exist unique scalars c_i such that

$$\widetilde{\psi} = \widetilde{\beta_0}\, c_0 + \cdots + \widetilde{\beta_r}\, c_r .$$

None of the coefficients c_i can be equal to zero, for otherwise a proper subset of $\{B_0, \cdots, B_r, U\}$ would be independent, contradicting our assumption about such proper subsets. If we now take $\beta_i = \widetilde{\beta_i}\, c_i$ for each i, we then have the desired relation $(\ddagger)$.□

Conversely, suppose that $(\ddagger)$ is satisfied. If we are given arbitrary homogeneous coordinates β_i' and ψ' for the points B_i and U, then there exist nonzero scalars a and q_i such that $\psi = \psi' a$ and $\beta_i = \beta_i' q_i$. The new homogeneous coordinate vectors satisfy a relation of the form

$$\psi' = \beta_0' q_0 a^{-1} + \cdots + \beta_r' q_r a^{-1}$$

and if $(\ddagger)$ is valid then all the coefficients on the right hand side must be equal to 1. In other words, we must have $b_i a^{-1} = 1$ for all i or equivalently $b_i = a$ for all i, which is exactly what we wanted to prove.■

Assume now that $(\ddagger)$ is valid, and let X be any point of the r-plane $L = B_0 \cdots B_r$. A set of homogeneous coordinates ξ for X is then a linear combination of the form $\xi = \sum_i \beta_i x_i$. Since ξ is defined up to multiplication by a scalar factor and the vectors β_j are defined up to multiplication by a common scalar factor, it follows that the coefficients x_i are also determined up to multiplication by a common scalar factor, and such an ordered $(r+1)$-tuple $(x_0, \cdots, x_r)$ of coefficients is called a set of *homogeneous coordinates for $X \in L$ relative to the projective coordinate system* $(B_0 \cdots B_r \,|\, U)$. It is frequently denoted by notation such as $\vec{X}(B_0 \cdots B_r \,|\, U)$. The set $\{B_0, \cdots, B_r\}$ is often called the *coordinate simplex* or *fundamental simplex*, the points B_i are said to be the *vertices* of this coordinate simplex, and the point U is often called the *unit point* because homogeneous coordinates for this point in the projective coordinate system are given by $(1, \cdots, 1)$.

The homogeneous coordinates given in the definition of projective space may be viewed as a special case of the preceding construction; specifically, if the unit vectors in $\mathbb{F}^{n+1,1}$ are given by $\mathbf{e}_i$, then the appropriate corrdinate simplex has vertices $\mathbf{e}_i \cdot \mathbb{F}$ and the corresponding unit point is $\mathbf{d} \cdot \mathbb{F}$, where $\mathbf{d} = \sum_i \mathbf{e}_i$. This is often called the *standard coordinate system.*

The next result describes the change in homogeneous coordinates which occurs if we switch from one projective coordinate system to another.

THEOREM VI.9. *Let* $(B_0 \ \cdots \ B_r \,|\, U)$ *and* $(B_0^* \ \cdots \ B_r^* \,|\, U^*)$ *be two projective coordinate systems for an* r*-plane in* $\mathbb{FP}^n$*, and let* X *be a point in this* r*-plane. Then the homogeneous coordinates* x_i *and* x_i^* *of* X *relative to these respective coordinate systems are related by the coefficients of an invertible matrix* $A = (\,a_{i,j}\,)$ *as follows:*

$$x_i^* \cdot \rho \;=\; \sum_{k=0}^{r} a_{i,k}\, x_k$$

Here ρ *is a nonzero scalar in* $\mathbb{F}$.

Proof. Suppose that the coordinate vectors are chosen as before so that $\psi = \sum_i \beta_i$ and $\psi' = \sum_i \beta_i'$. If ξ is a set of homogeneous coordinates for X, then homogeneous coordinates for ξ are defined by the two following two equations:

$$\xi \;=\; \sum_i \beta_i\, x_i \qquad\qquad \xi^* \;=\; \sum_i \beta_i^*\, x_i^*$$

Since the points lie in the same r-plane, we have

$$\beta_i^* \;=\; \sum_k \beta_k\, a_{i,k}$$

for sutiable scalars $a_{i,k}$, and the matrix A with these entries must be invertible because the set $\{B_0 \ \cdots \ B_r\}$ is independent. Straightforward calculation shows that

$$\xi \;=\; \sum_k \beta_k\, x_k \;+\; \sum_{i,k} \beta_i^*\, a_{i,k}\, x_k \;=\; \sum_i \beta_i^*\, x_i^*$$

which implies that $x_i^* = \sum_k a_{i,k}\, x_k$. These are the desired equations; we have added a factor ρ because the homogeneous coordinates are defined only up to a common factor.■

EXERCISES

1. Take the projective coordinate system on $\mathbb{RP}^3$ whose fundamental simplex points B_i have homogeneous coordinates

$$\beta_0 = \begin{pmatrix}1\\1\\0\\0\end{pmatrix}, \qquad \beta_1 = \begin{pmatrix}0\\0\\1\\1\end{pmatrix}, \qquad \beta_2 = \begin{pmatrix}0\\1\\1\\0\end{pmatrix}, \qquad \beta_3 = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

and whose unit point U has homogeneous coordinates

$$\psi = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}.$$

Find the homogeneous coordinates of the point A with respect to the system $(B_0\, B_1\, B_2\, B_3 \mid U)$ where ordinary homogeneous coordinates α for A are given below; there are two parts to this exercise corresponding to the two possibilities for A.

$$\alpha \;=\; \begin{pmatrix} 2 \\ 1 \\ 4 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ 1 \\ 3 \\ 0 \end{pmatrix}.$$

2. Let T be an invertible linear transformation on $\mathbb{F}^{n+1,1}$ with associated invertible matrix A, let T_* be the associated geometric symmetry of $\mathbb{FP}^n$, let $(B_0 \ \cdots \ B_n \mid U)$ define the standard homogeneous coordinate system, and let $X \in \mathbb{FP}^n$ have homogeneous coordinates given by $x_0, \ \cdots \ , x_n$. What are the homogeneous coordinates of X with respect to the coordinate system

$$\big(\, T_*(B_0) \ \cdots \ T_*(B_n) \mid T_*(U) \,\big) \ ?$$

3. Collineations

At the beginning of Section II.6, we noted that an appropriate notion of isomorphism for figures in Euclidean space is given by certain 1–1 correspondences with special properties. If one analyzes the situation further, it turns out that the relevant class of 1–1 correspondences is given by maps which extend to *isometries* of the Euclidean n-space $\mathbb{R}^n$. Specifically, these are 1–1 mappings T from $\mathbb{R}^n$ to itself with the following two properties:

(i) If $\mathbf{x}$ and $\mathbf{y}$ are distinct points in $\mathbb{R}^n$, then T satisfies the identity

$$d(\mathbf{x},\,\mathbf{y}) \;=\; d\big(\,T(\mathbf{x}),\,T(\mathbf{y})\,\big)\ ;$$

in other words, T preserves distances between points.

(ii) If $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are distinct points in $\mathbb{R}^n$, then T satisfies the identity

$$\mathsf{measure}(\angle\,\mathbf{xyz}) \;=\; \mathsf{measure}\big(\,\angle\,T(\mathbf{x})T(\mathbf{y})T(\mathbf{z})\,\big)\ ;$$

in other words, T preserves angle measurements.

Further information on such mappings and closely related issues can be found in the Addendum to Appendix A and the references cited there. For our purposes here, one important point is that one can describe all such isometries of $\mathbb{R}^n$ in a very simple and explicit manner. Specifically, every such isometry of $\mathbb{R}^n \cong \mathbb{R}^{n,1}$ is given by a mapping of the form $T(\mathbf{x}) \;=\; A\mathbf{x} + \mathbf{b}$, where $\mathbf{b} \in \mathbb{R}^n \cong \mathbb{R}^{n,1}$ and A is an $n \times n$ matrix which is *orthogonal*; the latter means that ${}^{\mathbf{T}}\!A$ is equal to A^{-1} or equivalently that the rows and columns of A define orthonormal sets of vectors. In this section we shall prove similar results for symmetries of projective spaces, showing that all geometric symmetries of $\mathbb{FP}^n$ are also given by some fairly basic constructions using linear algebra.

Frequently in this section we shall use the term **collineation** to denote an isomorphism from one n-dimensional incidence space to another (assuming $n \geq 2$). This name dates back to the 19^{th} century, and at the time collineations were the first types of incidence space isomorphisms to be considered abstractly.

Algebraic automorphisms and geometric symmetries

We have seen that every invertible $(n+1) \times (n+1)$ matrix A determines a collineation f_A of $\mathbb{FP}^n$ which is defined by the formula

$$f_A(\mathbf{x}\cdot\mathbb{F}) \;=\; A\mathbf{x}\cdot\mathbb{F}\ .$$

However, for many choices of $\mathbb{F}$ there are examples which do not have this form. In particular, if $\mathbb{F}$ is the complex numbers $\mathbb{C}$ and χ denotes the map on $\mathbb{F}^{n+1,1}$ which takes a column vector with entries z_j to the column vector whose entries are the complex conjugates[1] $\overline{z_j}$, then there is a well-defined collineation g_χ on $\mathbb{CP}^n$ such that

$$g_\chi(\mathbf{x}\cdot\mathbb{C}) \;=\; \chi(\mathbf{x})\cdot\mathbb{C}$$

that can also be defined, but it turns out that such a map is not equal to any of the maps f_A described previously. The proof that g_χ is a collineation depends upon the fact that complex conjugation is an automorphism; *i.e.*, we have $\overline{z_1 + z_2} \;=\; \overline{z_1} + \overline{z_2}$, $\overline{z_1 \cdot z_2} \;=\; \overline{z_1}\cdot\overline{z_2}$, and conjugation is a 1–1 correspondence to $\mathbb{C}$ to itself because this map is equal to its own inverse.

[1]Recall that if a complex number is given by $u + v\mathbf{i}$, where $\mathbf{i}^2 = -1$, then its conjugate is equal to $a - b\mathbf{i}$.

More generally, if $\mathbb{F}$ is an arbitrary skew-field and χ is an automorphism of $\mathbb{F}$, then one can construct a similar collineation g_χ on $\mathbb{FP}^n$ that is not expressible as f_A for some A. One major objective of this section is to prove that mappings of the form f_A and g_χ for the various choices of A and χ determine **all** collineations of $\mathbb{FP}^n$. In order to simplify the arguments, for the rest of this section we shall assume that the skew-field $\mathbb{F}$ is commutative; at the end of the section we shall discuss some aspects of the noncommutative case.

In fact, one of the most important prepreties of a collineation f from one coordinate projective space to another (perhaps over a different field) is that the collineation determines an isomorphism Φ_f of the underlying fields; if the two projective spaces are identical, this isomorphism becomes an automorphism. The first result of this section establishes the relationship between collineations and field isomorphisms.

THEOREM VI.10. *Let f be a collineation from the projective space $\mathbb{FP}^n$ to the projective space $\mathbb{EP}^n$, where $\mathbb{F}$ and $\mathbb{E}$ are fields and $n \geq 2$. Then there is an isomorphism*

$$\Phi_f : \mathbb{F} \longrightarrow \mathbb{E}$$

characterized by the equation

$$\Phi_f\big(\mathsf{XR}(Y_1,\, Y_2,\, Y_3,\, Y_4)\big) \;=\; \mathsf{XR}\big(\,\Phi_f(Y_1),\, \Phi_f(Y_2),\, \Phi_f(Y_3),\, \Phi_f(Y_4)\,\big)$$

where $Y_1,\ Y_2,\ Y_3,\ Y_4$ is an arbitrary sequence of collinear points such that the first three are distinct and $Y_4 \neq Y_2$.

There are three basic steps in the proof; namely, defining a map from $\mathbb{F}$ to $\mathbb{E}$ which dependes upon some choices, showing that such a map is an isomorphism of fields, and finally showing that the map is independent of the choices that were made at the first step. The second part of the proof uses the results from Section V.5, and the third part — which is by far the longest — relies heavily on results from Chapter V on cross ratios.

Proof of Theorem 10. *Construction of a mapping from $\mathbb{F}$ to $\mathbb{E}$.* Let X_0, X_1 and U be three distinct collinear points, and let $q \in \mathbb{F}$. Then there is a unique point $Q \in X_0X_1$ such that $Q \notin P_0$ and $\mathsf{XR}(X_0, X_1, U, Q) \;=\; q$. Define $\Phi_f(q) \;=\; \mathsf{XR}\big(f(X_0), f(X_1), f(U), f(Q)\big)$. Strictly speaking, one should write this as $\Phi_{f,X_0,X_1,U}$ to indicate that it depends upon the choices of X_0, X_1 and U.

We claim that *the map $\Phi_{f,X_0,X_1,U}$ defines an isomorphism from $\mathbb{F}$ to $\mathbb{E}$.* — Since the elements of $\mathbb{E}$ are in 1–1 correspondence with the elements of $f(X_0)f(X_1) - \{f(X_0)\}$ and f maps $X_0X_1 - \{X_0\}$ bijectively to $f(X_0)f(X_1) - \{f(X_0)\}$, it follows that $\Phi_{f,X_0,X_1,U}$ is 1–1 and onto. Furthermore, to see that the latter map is an isomorphism, take another line L through X_0, coplanar points Z_0 and D, and points $A,\, B \in X_0X_1$ as in Section V.3. Let $f(X_i) = X'_i$, $f[L] = L'$, $f(Z_0) - Z'_0$, $f(A) = A'$, $f(D) = D'$, and $f(B) = B'$. If X is any point constructed from the unprimed point as in Section V.5, let X' be the corresponding point constructed from the primed points. Since f is a collineation, it is easy to verify that $f(X) = X'$ for all point X constructed in Section V.5. In particular, $f(C) = C'$ and $f(K) = K'$. But the latter equalities combined with Theorem V.28 and V.29 imply that

$$\Phi_{f,X_0,X_1,U}(a+b) \;=\; \Phi_{f,X_0,X_1,U}(a) \;+\; \Phi_{f,X_0,X_1,U}(b)\ .$$

Proof that the map $\Phi_{f,X_0,X_1,U}$ depends only on the line M containing X_0, X_1 and U. — It suffices to show that

$$\Phi_{f,X_0,X_1,U}\big(\mathsf{XR}(A,B,C,D)\big) \;=\; \mathsf{XR}\big(\,f(A),f(B),f(C),f(D)\,\big)$$

for every quadruple of distinct points on X_0X_1. There are several cases to be considered.

CASE 1. None of the points A, B, C, D is X_0. If we choose homogeneous coordinates ξ_i for X_i and ψ for U such that $\xi_0+\xi_1=\psi$, then homogeneous coordinates α, β, γ and δ for A, B, C and D are given as follows:

$$\begin{aligned}\alpha &= \mathsf{XR}(X_0,X_1,U,A)\xi_0 \;+\; \xi_i\\ \beta &= \mathsf{XR}(X_0,X_1,U,B)\xi_0 \;+\; \xi_i\\ \gamma &= \mathsf{XR}(X_0,X_1,U,C)\xi_0 \;+\; \xi_i\\ \delta &= \mathsf{XR}(X_0,X_1,U,D)\xi_0 \;+\; \xi_i\end{aligned}$$

and therefore we have

$$\Phi_{f,X_0,X_1,U}\big(\mathsf{XR}(A,B,C,D)\big) \;=\; \mathsf{XR}\big(\,f(A),f(B),f(C),f(D)\,\big)$$

by the formula established in Theorem V.13 and the fact that $\Phi_{f,X_0,X_1,U}$ is an isomorphism of fields.

CASE 2. One of the points is X_0. We claim it suffices to consider the case $A=X_0$. For by Theorem V.12 there is a reordering $\big(\sigma(A),\sigma(B),\sigma(C),\sigma(D)\big)$ of (A,B,C,D) such that $\sigma(A)=X_0$ and

$$\mathsf{XR}\big(\sigma(A),\sigma(B),\sigma(C),\sigma(D)\big) \;=\; \mathsf{XR}(A,B,C,D)\ .$$

If the assertion is correct for quadruples whose first term is X_0, then

$$\Phi_{f,X_0,X_1,U}\Big(\,\big(\sigma(A),\sigma(B),\sigma(C),\sigma(D)\big)\Big) = \Big(\,\sigma\big(\,f(A)\big),\sigma\big(\,f(B)\big),\sigma\big(\,f(C)\big),\sigma\big(\,f(D)\big)\,\Big)\ .$$

Since the right hand side is equal to $\Phi_{f,X_0,X_1,U}\big(\mathsf{XR}(A,B,C,D)\big)$ and the right hand side is equal to $\mathsf{XR}\big(\,f(A),f(B),f(C),f(D)\big)$, the cases where X_0 is one of B, C or D follow.

By the preceding discussion, we might as well assume that $X_0=A$ in Case 2. The remainder of the argument for Case 2 splits into subcases depending upon whether X_1 is equal to one of the remaining points.

SUBCASE 2.1. Suppose that $A=X_0$ and $B=X_1$. Then by Theorem V.11 we have

$$\mathsf{XR}(A,B,C,D) \;=\; \frac{\mathsf{XR}(A,B,U,D)}{\mathsf{XR}(A,B,U,C)} \;=\; \frac{\mathsf{XR}(X_0,X_1,U,D)}{\mathsf{XR}(X_0,X_1,U,C)}\ .$$

Note that the cross ratio $\mathsf{XR}(A,B,U,C)$ is nonzero because $B\neq C$. The assertion in this case follows from the formula above and the fact that Φ is an automorphism.

SUBCASE 2.2. Suppose that $A=X_0$ and $C=X_1$. Then $\mathsf{XR}(A,B,C,D) \;=\; 1-\mathsf{XR}(A,C,B,D)$, and hence the assertion in this subcase follows from Subcase 2.1 and the fact that Φ is an automorphism.

SUBCASE 2.3. Suppose that $A=X_0$ but neither B nor C is equal to X_1. Let

$$\begin{aligned}b &= \mathsf{XR}(X_0,X_1,U,B)\\ c &= \mathsf{XR}(X_0,X_1,U,C)\\ d &= \mathsf{XR}(X_0,X_1,U,D)\end{aligned}$$

so that homogeneous coordinates β, ξ_0 and ξ_1 for the points B, X_0 and X_1 can be chosen such that $\beta = b\xi_0 + \xi_1$, and hence the corresponding homogeneous coordinates $\gamma = c\xi_0 + \xi_1$ for C satisfy

$$\gamma \;=\; c\xi_0 \;+\; \xi_1 \;=\; (c-b)\xi_0 \;+\; (b\xi_0 + \xi_1)\ .$$

Since $B \neq C$, it follows that $c - b = 0$. Therefore homogeneous coordinates δ for D are given by

$$\delta \;=\; d\xi_0 \;+\; \xi_1 \;=\; \frac{d-b}{c-b}(c-b)\xi_0 \;+\; (b\xi_0 + \xi_1)\ .$$

Therefore we have the identity

$$\mathsf{XR}(A,B,C,D) \;=\; \frac{d-b}{c-b}\ .$$

The assertion in this subcase follows from the above formula and the fact that Φ is an isomorphism. This concludes the proof that Φ only depends upon the line $L = X_0X_1$.

Proof that the isomorphism $\Phi_f = \Phi_{f,M}$ does not depend upon the choice of the line M. — Once again, there are two cases.

CASE 1. Suppose we are given two lines M and M' which have a point in common; we claim that $\Phi_{f,M} = \Phi_{f,M'}$. Let V be a point in the plane of M and M' which is not on either line. If $X \in M$, let $X' \in M' \cap VX$; then

$$f(X') \;\in\; f[M'] \;\cap\; f(V)f(X)$$

because f is a collineation. Thus two applications of Theorem 15 imply

$$\mathsf{XR}(A,B,C,D) \;=\; \mathsf{XR}(A',B',C',D')$$

$$\mathsf{XR}\big(\,f(A),f(B),f(C),f(D)\,\big) \;=\; \mathsf{XR}\big(\,f(A'),f(B'),f(C'),f(D')\,\big)\ .$$

On the other hand, we also have

$$\Phi_{f,M}\big(\mathsf{XR}(A,B,C,D)\big) \;=\; \mathsf{XR}\big(\,f(A),f(B),f(C),f(D)\,\big) \text{ and}$$

$$\Phi_{f,M'}\big(\mathsf{XR}(A',B',C',D')\big) \;=\; \mathsf{XR}\big(\,f(A'),f(B'),f(C'),f(D')\,\big)\ .$$

Since every element of $\mathbb{F}$ has the form $\mathsf{XR}(A,B,C,D)$ for suitable points, the equations above imply that $\Phi_{f,M} = \Phi_{f,M'}$.

CASE 2. Suppose we are given two lines M and M' which have no points in common; we claim that $\Phi_{f,M} = \Phi_{f,M'}$. Let M'' be a line joining one point in M to one point in M'. Then two applications of the first case imply that $\Phi_{f,M} = \Phi_{f,M''} = \Phi_{f,M'}$.■

The characterization of Φ_f in terms of the cross ratio implies some useful properties of the construction Φ which sends a collineation $\mathbb{FP}^n \to \mathbb{EP}^n$ to the field isomorphism $\Phi_f : \mathbb{F} \to \mathbb{E}$.

THEOREM VI.11. *If $f : P \to P'$ and $f' : P' \to P''$ are collineations of coordinate projective n-spaces (where $n \geq 2$), then $\Phi_{gf} = \Phi_g \circ \Phi_f$. If $f : P \to P$ is the identity, then Φ_f is the identity on the underlying field. Finally, if $g : P' \to P$ is equal to f^{-1}, then $\Phi_g = (\Phi_f)^{-1}$.*

Proof. If f is the identity, then we have

$$\Phi_f\big(\mathsf{XR}(A,B,C,D)\big) \;=\; \mathsf{XR}\big(\,f(A),f(B),f(C),f(D)\,\big) \;=\; \mathsf{XR}(A,B,C,D)$$

because $f(X) = X$ for all X. If g and f are collineations then

$$\Phi_g \circ \Phi_f\big(\mathsf{XR}(A,B,C,D)\big) \;=\; \Phi_g\Big(\ \mathsf{XR}\big(\,f(A),f(B),f(C),f(D)\,\big)\ \Big) \;=$$

$$\mathsf{XR}\big(\, g\circ f(A), g\circ f(B), g\circ f(C), g\circ f(D)\,\big) \;=\; \Phi_{gf}\big(\,\mathsf{XR}(A,B,C,D)\,\big)\ .$$

To prove that $\Phi_{f^{-1}} = (\Phi_f)^{-1}$, note that $f\circ f^{-1} =$ identity and $f^{-1}\circ f =$ identity combine with the first two identities to show that the composites $\Phi_{f^{-1}}\circ\Phi_f$ and $\Phi_f\circ\Phi_{f^{-1}}$ are both identity maps, and these identities imply that $\Phi_{f^{-1}} = (\Phi_f)^{-1}$.■

Collineations of $\mathbb{FP}^1$

Of course, an incidence-theoretic definition of collineations for coordinate projective lines is meaningless. However, if $1+1 \neq 0$ in $\mathbb{F}$, then as in Section V.4 it is possible to define collineations of $\mathbb{FP}^1$ as 1–1 correspondences which preserve harmonic quadruples. With this definition, an analog of Theorem 10 is valid. Details appear on pages 85–87 of the book by Bumcrot listed in the bibliography (this is related to the discussion of von Staudt's Theorem at the end of Section V.4).

Examples

We have already noted that every invertible $(n+1)\times(n+1)$ matrix A over $\mathbb{F}$ defines a geometric symmetry f_A of $\mathbb{FP}^n$, and by a straightforward extension of Exercise V.2.10 the mapping f_A preserves cross ratios; therefore, the automorphism Φ_{f_A} is the identity. On the other hand, if χ is an automorphism of $\mathbb{F}$ as above and g_χ is defined as at the beginning of this section, then for all distinct collinear points $A,\ B,\ C,\ D$ in $\mathbb{FP}^n$ we have

$$\chi\big(\,\mathsf{XR}(A,B,C,D)\,\big) \;=\; \mathsf{XR}\big(\, g_\chi(A), g_\chi(B), g_\chi(C), g_\chi(D)\,\big)$$

and therefore $\Phi_{g_\chi} = \chi$. In particular, the latter implies the following:

> *For every field* $\mathbb{F}$*, every automorphism* χ *of* $\mathbb{F}$*, and every* $n > 0$*, there is a collineation* g *from* $\mathbb{FP}^n$ *to itself such that* $\Phi_g = \chi$.■

Later in the section we shall prove a much stronger result of this type.

The Fundamental Theorem of Projective Geometry

Before stating and proving this result, we need to state and prove some variants of standard results from linear algebra. Let V and W be vector spaces over a field $\mathbb{F}$, and let α be an automorphism of $\mathbb{F}$. A mapping $T : V \to V$ is said to be an *α-semilinear transformation* if it satisfies the following conditions:

(1) $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x})+T(\mathbf{y})$ for all $\mathbf{x},\ \mathbf{y}\in V$.

(2) $T(c\mathbf{x}+\mathbf{y}) = \alpha(c)T(\mathbf{x})$ for all $\mathbf{x}\in V$ and $c\in\mathbb{F}$.

If α is the identity mapping, this reduces to the usual definition of a linear transformation.

THEOREM VI.12. *Let* $V,\ W,\ \mathbb{F}, \alpha$ *be as above. If* $\mathbf{v}_1,\ \cdots, \mathbf{v}_n$ *is a basis for* V *and* $\mathbf{w}_1,\ \cdots, \mathbf{w}_n \in W$*, then there is a unique* α*-semilinear transformation* $T : V \to W$ *such that* $T(\mathbf{v}_i) = \mathbf{w}_i$ *for all* i*.*

Proof. Suppose that $v \in V$ and that T and S are α-semilinear transformations from V to W satisfying the conditions of the theorem. Write $\mathbf{v}$ as a linear combination $\sum_i c_i\mathbf{v}_i$. Then we have

$$T(\mathbf{v}) \;=\; T\left(\sum_i c_i\mathbf{v}_i\right) \;=\; \sum_i \alpha(c_i)T(\mathbf{v}_i) \;=\; \sum_i \alpha(c_i)\mathbf{w}_i \;=$$

$$\sum_i \alpha(c_i)S(\mathbf{v}_i) \;=\; S\left(\sum_i c_i\mathbf{v}_i\right) \;=\; S(\mathbf{v})$$

and hence $S = T$. Conversely, if $\mathbf{v}$ is given as above, then $T(\mathbf{v}) \;=\; \sum_\mathbf{i} \alpha(\mathbf{c_i})\mathbf{w_i}$ is a well-defined α-semilinear transformation.■

This result has the following basic consequence:

Theorem VI.13. *In the setting above, the mapping T is $1-1$ and onto if and only if the vectors $\mathbf{w}_1, \cdots, \mathbf{w}_n$ form a basis for W. In this case the inverse map T^{-1} is an α^{-1}-semilinear transformation.*

Proof. Since the image of T is contained in the subspace spanned by the vectors $\mathbf{w}_i$, it follows that T cannot be onto if these vectors do not span W. Conversely, suppose that these vectors do form a basis. Then by the previous result there is an α^{-1}-semilinear transformation $S : W \to V$ such that $S(\mathbf{w}_i) \;=\; \mathbf{v}_i$ for all i. It follows that $S\circ T$ is an $\alpha^{-1}\circ\alpha$-semilinear (hence linear) transformation from V to itself which sends $\mathbf{v}_i$ to $\mathbf{v}_i$ for all i, and hence $S\circ T$. Reversing the roles of V and W and also the roles of S and T in this argument, we conclude similarly that $T\circ S$ is the identity. Therefore the α-semilinear map S is an inverse to T and the latter is 1–1 and onto.■

If $\mathbb{F}$ and n are as in Theorem 12 and T is an invertible α-semilinear transformation from $\mathbb{F}^{n+1,1}$ to itself, then as in Section 4.3 there is a collineation f_T from $\mathbb{FP}^n$ to itself defined by

$$f_T(X) \;=\; T(\xi)\cdot\mathbb{F}$$

where ξ is an arbitrary set of homogeneous coordinates for X; this does not depend upon the choice of homogeneous coordinates, for if $\xi' = c\xi$ is another set of homogeneous coordinates for X we have

$$T(c\xi)\cdot\mathbb{F} \;=\; \alpha(c)\cdot T(\xi)\cdot\mathbb{F} \;=\; T(\xi)\cdot\mathbb{F}\,.$$

The proof that this map defines a collineation proceeds exactly as in the case of linear transformations, the only change being the need to substitute $T(c\mathbf{v}) \;=\; \alpha(c)\cdot T(\mathbf{v})$ in place of $T(c\mathbf{v}) \;=\; c\cdot T(\mathbf{v})$ when the latter appears.

The Fundamental Theorem of Projective Geometry is a converse to the preceding construction, and it shows that every collineation of $\mathbb{FP}^n$ to itself has the form f_T for a suitably chosen invertible α-linear mapping T from $\mathbb{F}^{n+1,1}$ to itself.

Theorem VI.14. (Fundamental Theorem of Projective Geometry) *Let $\{X_0, \cdots, X_n, A\}$ and $\{Y_0, \cdots, Y_n, B\}$ be two sets of $(n+2)$ points in $\mathbb{FP}^n$ (where $n \geq 2$) such that no proper subset of either is dependent, and let χ be an automorphism of $\mathbb{F}$. Then there is a unique collineation f of $\mathbb{FP}^n$ to itself satisfying the following conditions:*

(i) $f(X_i) \;=\; Y_i$ *for* $0 \leq i \leq n$.

(ii) $f(A) \;=\; B$.

(iii) $\Phi_f \;=\; \chi$.

The theorem (with the proof given here) is also valid if $n = 1$ and $1 + 1 \neq 0$ in $\mathbb{F}$, provided collineations of $\mathbb{FP}^1$ are defined in the previously described manner (*i.e.*, preserving harmonic quadruples).

Proof. EXISTENCE. According to Theorem 8 we can choose homogeneous coordinates ξ for X_i, η_i for Y_i, α for A, and β for B so that $\alpha = \sum_i \xi_i$ and $\beta = \sum_i \eta_i$. The hypotheses imply that the vectors ξ_i and η_i form bases for $\mathbb{F}^{n+1,1}$, and therefore there is an invertible α-semilinear transformation of the latter such that $T(\xi_i) = \eta_i$ for all i. Then f_T is a collineation of $\mathbb{FP}^n$ sending X_i to Y_i and A to B. In order to compute the automorphism induced by f, let $Q_1,\ Q_2,\ Q_3,\ Q_4$ be distinct collinear points with homogeneous coordinates θ_i for Q_i chosen such that $\theta_3 = \theta_1 + \theta_2$ and $\theta_4 = q\theta_1 + \theta_2$, where $q = \mathsf{XR}(Q_1, Q_2, Q_3, Q_4)$. We then have $T(\theta_3) = {}_T(\theta_1) +_T (\theta_2)$ and $T(\theta_4) = \chi(q) \cdot T(\theta_1) + T(\theta_2)$, so that

$$\chi\Big(\mathsf{XR}\big(Q_1, Q_2, Q_3, Q_4\big) \Big) = \mathsf{XR}\big(f_T(Q_1), f_T(Q_2), f_T(Q_3), f_T(Q_4) \big) .$$

It follows that $\Phi_{f_T} = \chi$.$\square$

UNIQUENESS. Suppose that f and g are collineations of $\mathbb{FP}^n$ which satisfy $f(X_i) = g(X_i) = Y_i$ for $0 \leq i \leq n$, $f(A) = g(A) = B$, and $\Phi_f = \Phi_g = \chi$. Then $h = g^{-1} \circ f$ is a collineation which satisfies $h(X_i) = X_i$ for $0 \leq i \leq n$, $h(A) = A$, and Φ_h is the identity. If suffices to show that a collineation h satisfying these conditions must be the identity.

Let h be a collineation such that Φ_h is the identity, and suppose that h leaves three distinct points on a line fixed; we claim that h leaves every point on the line fixed. To see this, suppose that X_1, X_2 and X_3 are distinct collinear points such that $h(X_i) = X_i$ for all i, and let $Y \in X_1X_2$. Then we have

$$\mathsf{XR}\big(X_1, X_2, X_3, h(Y) \big) = \Phi_h\big(\mathsf{XR}(X_1, X_2, X_3, Y) \big) = \mathsf{XR}(X_1, X_2, X_3, Y)$$

and hence $h(Y) = Y$ by Theorem V.10.

Now assume that h satisfies the conditions in the first paragraph of this argument. We shall prove, by induction on r, that h fixes every point in the r-plane $X_0 \cdots X_r$. The statement is trivially true for $r = 0$, so assume that it is true for $r - 1$, where $r \geq 1$. By the dimension formula, the intersection of the subspaces $X_0 \cdots X_r$ and $AX_{r+1} \cdots X_n$ is a point which we shall call B_r. In fact, homogeneous coordinates β_r for B_r are given by

$$\beta_r = \xi_0 + \cdots + \xi_r$$

because the right hand side is set of homogeneous coordinates for a point in the intersection. Since $h(X_i) = X_i$ and $h(A) = A$, it follows that h maps the subspaces $X_0 \cdots X_r$ and $AX_{r+1} \cdots X_n$ into themselves. Thus the intersection of these subspaces (namely, the one point set $\{B_r\}$) must be mapped into itself and hence $h(B_r) = B_r$.

We claim that h fixes every point on the line X_rB_r fixed. By hypothesis $h(X_r) = X_r$, and by the preceding paragraph $h(B_r) = B_r$. Hence h maps X_rB_r into itself. Since X_rB_r and $X_0 \cdots X_{r-1}$ are both contained in $X_0 \cdots X_r$, the dimension formula implies that they intersect in a point W. Since $W \in X_0 \cdots X_{r-1}$, the induction hypothesis implies that $h(W) = W$. Homogeneous coordinates ω for W are given by

$$\omega = \beta_r - \xi_r = \xi_0 + \cdots + \xi_{r-1}$$

and hence the points X_r, V_r and W are distinct collinear points. Since each is left fixed by h, it follows that every other point in X_rB_r is also left fixed by h.

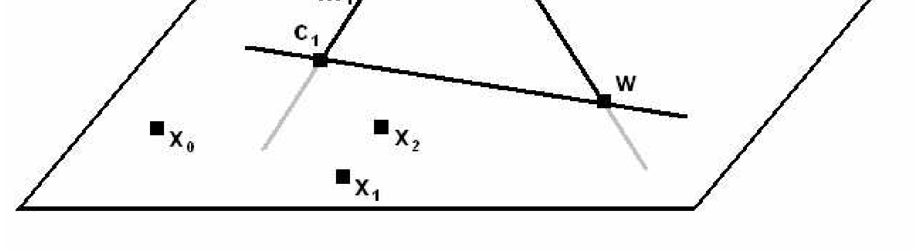


Figure VI.1

$(r = n = 3)$

Now let Z be any point of $X_0 \cdots X_r$ not on either $X_0 \cdots X_{r-1}$ or X_rB_r. We claim that $h(Z) = Z$. Since X_rB_r and $X_0 \cdots X_{r-1}$ are both contained in $X_0 \cdots X_r$, the dimension formula implies that $X_0 \cdots X_{r-1}$ and the plane ZX_rB_r intersect in a line we shall call L. The assumption that $Z \notin X_0 \cdots X_{r-1}$ implies that $Z \notin L$.

Let M_1 and M_2 be two distinct lines in ZX_rB_r containing Z; since there are at least three lines in the plane ZX_rB_r containing Z, we may choose M_1 and M_2 such that neither line contains the point B_r in which L meets X_rB_r; in particular, this means that the intersections of M_i with L and X_rB_r are distinct points.

Let $S_i \in M_i \cap L$, and let $D_i \in M_2 \cap X_rB_r$ (here $i = 1$ or 2). Then $h(C_i) = C_i$ and $h(D_i) = D_i$. Since the intersections of M_i with L and X_rB_r are distinct points, it follows that h leaves two distinct points of M_i fixed and hence h maps each line M_i into itself. Therefore it also follows that h maps $M_1 \cap M_2 = \{Z\}$ into itself, so that $h(Z) = Z$.

The preceding argument shows that h leaves every point of $X_0 \cdots X_r$ fixed, completing the inductive step of the argument. Therefore, by induction we conclude that h is the identity on $\mathbb{FP}^n = X_0 \cdots X_n$.■

One immediate consequence of the Fundamental Theorem is particularly worth stating at this point:

THEOREM VI.15. *Let f be a collineation of $\mathbb{FP}^n$, and let $\Phi_f = \alpha$. Then there is an invertible α-semilinear transformation T of $\mathbb{F}^{n+1,1}$ such that if $X \in \mathbb{FP}^n$ and ξ is a set of homogeneous coordinates for X then $f(X) = T(\xi) \cdot \mathbb{F}$.*

Proof. Let $\{X_0, \cdots, X_n, A\}$ be a set of of $(n+2)$ points in $\mathbb{FP}^n$ such that no proper subset is dependent. By the proof of existence in the Fundamental Theorem there is an invertible α-semilinear transformation T such that the associated collineation f_T satisfies the following conditions:

(*i*) $f(X_i) = f_T(X_i)$ for $0 \leq i \leq n$.

(*ii*) $f(A) = f_T(A)$.

Also, by construction the maps f and f_T determine the same automorphism of $\mathbb{F}$. We may now apply the uniqueness portion of the Fundamental Theorem to conclude that $f = f_T$.■

Definition. A collineation f of $\mathbb{FP}^n$ is *projective* if the associated automorphism Φ_f is the identity. Theorem 11 implies that the set of projective collineations is a subgroup — in fact, a normal subgroup — of the collineation group, and by the previous construction of examples we know that the quotient of the collineation group by the subgroup of projective collineations is equal to the automorphism group of $\mathbb{F}$. Further information along these lines is discussed in Exercise 9 below.■

Special cases

We conclude this section with some remarks on collineations if $\mathbb{F}$ is the real or complex numbers.

THEOREM VI.16. *For each $n \geq 2$, every collineation of real projective n-space $\mathbb{RP}^n$ is projective.*

By the previous results of this section, the proof of Theorem 15 reduces to showing the following:

THEOREM VI.17. *The only automorphism of the real numbers is the identity map.*

Proof(s). If χ is an isomorphism of $\mathbb{R}$, then $\chi(0) = 0$ and $\chi(0) = 1$. Suppose $\chi(r) = r$ for $r \geq 1$. Then $\chi(r+1) = \chi(r) + \chi(1) = r+1$, and hence χ agrees with the identity on all nonnegative integers. If k is a negative integer and $k = -m$, then

$$\chi(k) \ = \ \chi(-m) \ = \ -\chi(m) \ = \ -m \ = \ k$$

and hence χ is the identity on integers. If r is a rational number, write $r = m/n$ where m is an integer and n is a positive integer. Then $n = rm$ implies that

$$m \ = \ \chi(m) \ = \ \chi(nr) \ = \ \chi(n) \cdot \chi(r) \ = \ n \cdot \chi(r)$$

which implies that $\chi(r) = m/n = r$, and hence we see that χ fixes every rational number.

Suppose now that x is an arbitrary nonnegative real number. We claim that $\chi(x) \geq 0$. Recall that $x \geq 0$ if and only if $x = y^2$ for some y. Therefore $x \geq 0$ and $x = y^2$ imply that $\chi(x) = \chi(y)^2 \geq 0$. Similarly, if a and b are real numbers such that $a \geq b$, then

$$\chi(a) \ - \ \chi(b) \ = \ \chi(a-b) \ \geq \ 0$$

implies that $\chi(a) \geq \chi(b)$. Since χ is 1–1 it also follows that $a > b$ implies $\chi(a) > \chi(b)$.

Finally, suppose that we have an element $r \in \mathbb{R}$ such that $\chi(r) \neq r$. If $\chi(r) < r$, then there is a rational number q such that $\chi(r) < q < r$. But this implies $\chi(r) < \chi(q) = q$, and this contradicts the conclusion $\chi(r) > \chi(q)$ which follows from the previous paragraph. Therefore $\chi(r) < r$ is impossible, so that $\chi(r) \geq r$ for all real numbers r.

Now χ^{-1} is also an automorphism of $\mathbb{R}$, and if we apply the previous reasoning to this automorphism we conclude that $\chi^{-1}(r) \geq r$ for all r. Since we had previously shown that automorphisms are strictly increasing functions, if we apply χ to the previous inequality we obtain

$$r \ = \ \chi^\circ\chi^{-1}(r) \ \geq \ \chi(r)$$

and if we combine this with the final inequality of the preceding paragraph we conclude that $\chi(r) = r$ for all real numbers r.■

The analog of Theorem 16 does not hold for the complex numbers. In particular, at the beginning of this section we showed that the map g_χ of $\mathbb{CP}^n$ given by conjugating homogeneous coordinates is a collineation that is not projective. Further information on automorphisms of the complex numbers and their applications to projective geometry appears in Appendix D.

EXERCISES

In the problems below, assume that $\mathbb{F}$ is a field and χ is an automorphism of $\mathbb{F}$.

1. Let A be an invertible $n \times n$ matrix over $\mathbb{F}$, and let f_A be the projective collineation of $\mathbb{FP}^n$ defined by A (in other words, if ξ are homogeneous coordinates for X, then $A\xi \cdot \mathbb{F}$ are homogeneous coordinates for $f_A(X)$). If H is a hyperplane in $\mathbb{FP}^n$ with homogeneous coordinates θ, prove that $T[H]$ has homogeneous coordinates $\theta \cdot A^{-1}$ (compare Exercise V.1.5).

2. In the notation of Exercise 1, suppose that a collineation T is defined such that if ξ are homogeneous coordinates for X, then $A\chi(\xi)$ are homogeneous coordinates for $T(X)$. Express homogeneous coordinates for $T[H]$ in terms of θ, A^{-1} and χ. You may use the product formula $\chi(A \cdot B) \;=\; \chi(A) \cdot \chi(B)$ for matrix multiplication. Also, recall that $\chi(0) \;=\; 0$.

3. Suppose that f is a collineation of $\mathbb{FP}^n$ with induced automorphism Φ_f, and suppose that $H_1,\ H_2,\ H_3,\ H_4$ are distinct hyperplanes containing a common $(n-2)$-plane. Prove that the cross ratio formula

$$\Phi_f\big(\mathsf{XR}(H_1, H_2, H_3, H_4)\big) \;=\; \mathsf{XR}\big(f[H_1], f[H_2], f[H_3], f[H_4]\big)$$

holds without using Exercise 2.

4. Suppose that T is an invertible χ-semilinear transformation of $\mathbb{F}^{n+1,1}$ where $n \geq 1$ such that the associated collineation f_T of $\mathbb{FP}^n$ is the identity. Prove that T is a scalar multiple of the identity. [*Hint:* By assumption, for each nonzero vector $\mathbf{x}$ there is a nonzero scalar $c_{\mathbf{x}}$ such that $T(\mathbf{x}) \;=\; c_{\mathbf{x}} \cdot \mathbf{x}$. If $c_{\mathbf{x}} \neq c_{\mathbf{y}}$, explain why $\mathbf{x}$ and $\mathbf{y}$ must be linearly independent. Consider $T(\mathbf{x}+\mathbf{y})$ in this case.]

5. (a) Let T be an invertible χ-semilinear transformation of $\mathbb{F}^n$ where $n \geq 1$, and let $\mathbf{z} \in \mathbb{F}^n$. Show that

$$G(\mathbf{x}) \;=\; T(\mathbf{x}) \;+\; \mathbf{z}$$

is a geometric symmetry of the affine incidence n-space $\mathbb{F}^n$. [*Hint:* Compare this statement to the examples following Theorem II.39.]

(b) Prove that G extends to a collineation g of $\mathbb{FP}^n$ for which $\Phi_g \;=\; \chi$; in other words, we have $g \circ \mathrm{J} \;=\; \mathrm{J} \circ G$ on $\mathbb{F}^n$. [*Hint:* Compare Exercise IV.4.14.]

(c) If $n \geq 2$, prove that every geometric symmetry f of $\mathbb{F}^n$ is given by a transformation of the type described in (a). [*Hint:* By Exercise 2 at the end of Chapter IV, the map f extends to a collineation g of $\mathbb{FP}^n$. Since the collineation leaves the hyperplane at infinity fixed, certain entries of an $(n+1) \times (n+1)$ matrix inducing g must vanish. But this implies the matrix has the form of one constructible by (b).]

(d) Determine whether $\mathrm{Aff}(\mathbb{F}^n)$ is the entire group of geometric symmetries of $\mathbb{F}^n$ when $\mathbb{F}$ is the real and complex numbers respectively.

6. Suppose that A is an invertible $m \times m$ matrix over a field $\mathbb{F}$ such that $1+1 \neq 0$ in $\mathbb{F}$. Prove that $\mathbb{F}^{m,1}$ contains two vector subspaces W_+ and W_- with the following properties:

(i) $A\mathbf{x} \;=\; \mathbf{x}$ if $\mathbf{x} \in W_+$ and $A\mathbf{x} \;=\; -\mathbf{x}$ if $\mathbf{x} \in W_-$.

(ii) $W_+ \;+\; W_- \;=\; \mathbb{F}^{m,1}$ and $W_+ \;\cap\; W_- \;=\; \{\mathbf{0}\}$.

[*Hint:* Let $W_\pm$ be the image of $A \pm I$. This yields the first part. To prove the rest, use the identity

$$I \;=\; \tfrac{1}{2}(A+I) \;-\; \tfrac{1}{2}(A-I)\ .]$$

Definition. An *involution* of $\mathbb{FP}^n$ is a collineation f such that $f \circ f$ is the identity but f itself is not the identity. If $f(X) = X$, then X is called a **fixed point** of the involution.

7. (a) Let T be an invertible χ-semilinear transformation of $\mathbb{F}^{n+1,1}$ such that the induced collineation f_T of $\mathbb{FP}^n$ is an involution. Prove that T^2 is a scalar multiple of the identity. [*Hint:* Use Exercise 4.]

(b) Suppose that T is an involution of $\mathbb{RP}^n$. Prove that T is induced by an invertible $(n+1) \times (n+1)$ matrix A such that $A^2 = \pm 1$.

(c) In the previous part, prove that T has no fixed points if $A^2 = -I$. Using Exercise 6, prove that T has fixed points if $A^2 = 1$. [*Hint:* For the first part, suppose that X is a fixed point with homogeneous coordinates ξ such that $A \cdot \xi = c \cdot \xi$ for some real number c. However, $A^2 = -I$ implies that $c^2 = -1$.] — NOTATION. An involution is called *elliptic* if no fixed points exist and *hyperbolic* if fixed points exist.

(d) Using Exercise 6, prove that the fixed point set of a hyperbolic involution of $\mathbb{RP}^n$ has the form $Q_1 \cup Q_2$, where Q_1 and Q_2 are disjoint n_1- and n_2-planes and $n_1 + n_2 + 1 = n$.

8. Suppose that $A \neq B$, and that A and B are the only two points of the line AB left fixed by an involution f of $\mathbb{RP}^n$. Prove that $\mathsf{XR}\big(A, B, C, f(C)\big) = -1$ for all points C on $AB - \{A, B\}$. [*Hint:* Find an equation relating $\mathsf{XR}\big(A, B, C, f(C)\big)$ and $\mathsf{XR}\big(A, B, f(C), C\big)$.]

9. Let $\textsc{Coll}\,(\mathbb{FP}^n)$ denote the group of all collineations of $\mathbb{FP}^n$, let $\mathrm{Aut}(\mathbb{F})$ denote the group of (field) automorphisms of $\mathbb{F}$, and let $\Phi : \textsc{Coll}\,(\mathbb{FP}^n) \to \mathrm{Aut}(\mathbb{F})$ denote the homomorphism given by Theorem VI.10.

(a) Why is the kernel of Φ the group $\textsc{Proj}\,(\mathbb{FP}^n)$ of all projective collineations, and why does this imply that the latter is a normal subgroup of $\textsc{Coll}\,(\mathbb{FP}^n)$?

(b) Show that $\textsc{Coll}\,(\mathbb{FP}^n)$ contains a subgroup Γ isomorphic to $\mathrm{Aut}(\mathbb{F})$ such that the restricted homomorphism $\Phi|\Gamma$ is an isomorphism, and using this prove that every element of $\textsc{Coll}\,(\mathbb{FP}^n)$ is expressible as a product of an element in $\textsc{Proj}\,(\mathbb{FP}^n)$ with an element in Γ. [*Hint:* Look at the set of all collineations of the form g_χ constructed at the top of the second page of this section, where $\chi \in \mathrm{Aut}(\mathbb{F})$, and show that the set of all such collineations forms a subgroup of $\textsc{Coll}\,(\mathbb{FP}^n)$ which is isomorphic to $\mathrm{Aut}(\mathbb{F})$.]

(c) Suppose that A is an invertible $(n+1) \times (n+1)$ matrix over $\mathbb{F}$ and χ is an automorphism of $\mathbb{F}$, and let f_A and g_χ be the collineations of $\mathbb{FP}^n$ defined at the beginning of this section. By the previous parts of this exercise and Theorem 15, we know that $g_\chi \circ f_A \circ (g_\chi)^{-1}$ has the form f_B for some invertible $(n+1) \times (n+1)$ matrix B over $\mathbb{F}$. Prove that we can take B to be the matrix $\chi(A)$ obtained by applying χ to each entry of A. [*Note:* As usual, if two invertible matrices are nonzero scalar multiples of each other then they define the same projective collineation, and in particular we know that $f_B = f_{cB}$ for all nonzero scalars c; this is why we say that we take B to be equal to $\chi(A)$ and not that B is equal to $\chi(A)$.]

10. Let $\mathbb{F}$ be a field, let $0 < r \leq n$ where $n \geq 2$, let Q be an r-plane in $\mathbb{FP}^n$. Let $\{X_0, \cdots, X_r, A\}$ and $\{Y_0, \cdots, Y_r, B\}$ be two sets of $(r+2)$ points in Q such that no proper subset of either is dependent. Then there is a projective collineation f of $\mathbb{FP}^n$ to itself such that $f(X_i) = Y_i$ for $0 \leq i \leq r$ and $f(A) = B$. [*Hint:* Let W be the vector subspace of $\mathbb{F}^{n+1,1}$ such that $Q = \mathcal{S}_1(W)$, define an invertible linear transformation G on W which passes to a projective collineation of Q with the required properties as in the proof of the Fundamental Theorem, extend G to an invertible linear transformation $\overline{G}$ of $\mathbb{F}^{n+1,1}$, and consider the projective collineation associated to $\overline{G}$.]

4. Order and separation

All of the analytic projective geometry done up to this point is valid for an arbitrary $\mathbb{F}$ for which $1+1 \neq 0$. Certainly one would expect that real projective spaces have many properties not shared by other coordinate projective spaces just as the field of real numbers has many properties not shared by other fields. The distinguishing features of the real numbers are that it is an ordered field and is complete with respect to this ordering. In this section we shall discuss some properties of projective spaces over arbitrary ordered fields and mention properties that uniquely characterize real projective spaces.

Given points $\mathbf{u},\ \mathbf{v} \in \mathbb{R}^n$, let $d(\mathbf{u},\mathbf{v})$ be the distance from $\mathbf{u}$ to $\mathbf{v}$. One characterization of the statement

$$\mathbf{y} \textit{ is between } \mathbf{x} \textit{ and } \mathbf{z}$$

is that it holds if and only if $d(\mathbf{x},\mathbf{z}) \ = \ d(\mathbf{x},\mathbf{y})+d(\mathbf{y},\mathbf{z})$. Another more algebraic characterization follows immediately from this.

THEOREM VI.18. *If $\mathbf{x},\ \mathbf{y},\ \mathbf{z} \in \mathbb{R}^n$ are distinct points, then $d(\mathbf{x},\mathbf{z}) \ = \ d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})$ holds if and only if $\mathbf{y} \ = \ t\mathbf{x} + (1-t)\mathbf{z}$ for some t satisfying $0 < t < 1$.*

Proof. Recall that $d(\mathbf{u},\mathbf{v})$ is the square root of

$$(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v}) \ = \ |\mathbf{u}-\mathbf{v}|^2 \ .$$

The proof of the Triangle Inequality for inner (or dot) products is a consequence of the Cauchy-Schwarz inequality

$$(\mathbf{x}-\mathbf{y})\cdot(\mathbf{y}-\mathbf{z}) \ \leq \ \big|(\mathbf{x}-\mathbf{y})\cdot(\mathbf{y}-\mathbf{z})\big| \ \leq \ |\mathbf{x}-\mathbf{y}|\cdot|\mathbf{y}-\mathbf{z}|$$

and equality holds in the Triangle Inequality if and only if the end terms of the Cauchy-Schwarz inequality are equal.[2] However, the Cauchy-Schwarz inequality states that the middle term and right hand term are equal if and only if $\mathbf{x}-\mathbf{y}$ and $\mathbf{y}-\mathbf{z}$ are linearly dependent. Since both are nonzero, this means that $(\mathbf{y}-\mathbf{z}) \ = \ k(\mathbf{x}-\mathbf{y})$ for some $k \neq 0$. On the other hand, the left and right hand terms are equal if and only if both are nonnegative. Consequently, if the end terms are equal, then $(\mathbf{y}-\mathbf{z}) \ = \ k(\mathbf{x}-\mathbf{y})$ and also $k|\mathbf{x}-\mathbf{y}| \geq 0$. This implies that k must be positive. Conversely, if $k > 0$ then the end terms of the Cauchy-Schwarz inequality are equal.

Thus $d(\mathbf{x},\mathbf{z}) \ = \ d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})$ if and only if $\mathbf{y}-\mathbf{z}$ is a positive multiple of $\mathbf{x}-\mathbf{y}$. But if $\mathbf{y}-\mathbf{z} \ = \ k(\mathbf{x}-\mathbf{y})$, then

$$\mathbf{y} \ = \ \frac{k}{k+1}\,\mathbf{x} \ + \ \frac{1}{k+1}\,\mathbf{z} \ .$$

Since $k > 0$ implies

$$0 \ < \ \frac{k}{k+1} \ < \ 1$$

it follows that if $d(\mathbf{x},\mathbf{z}) \ = \ d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})$ then $\mathbf{y} \ = \ t\mathbf{x} + (1-t)\mathbf{z}$ for some t satisfying $0 < t < 1$.

Conversely, if $\mathbf{y} \ = \ t\mathbf{x} + (1-t)\mathbf{z}$ for some t satisfying $0 < t < 1$, then

$$\mathbf{y}-\mathbf{z} \ = \ \frac{t}{1-t}\,(\mathbf{x}-\mathbf{y}) \ .$$

[2]See pp. 177–178 of Birkhoff and MacLane or pp. 277–278 of Hoffman and Kunze for further details.

Since $t/(1-t)$ is positive if $0 < t < 1$, it follows that $d(\mathbf{x},\mathbf{z}) \; = \; d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})$.■

With this motivation, we define betweenness for arbitrary vector spaces over arbitrary ordered fields.

Definition. Let $\mathbb{F}$ be an ordered field, let V be a vector space over $\mathbb{F}$, and let $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ be distinct points of V. We shall say that $\mathbf{y}$ is **between** $\mathbf{x}$ and $\mathbf{z}$, written $\boldsymbol{\mathcal{B}}(\mathbf{x},\mathbf{y},\mathbf{z})$, if there is some $t \in \mathbb{F}$ such that $0 < t < 1$ and $\mathbf{y} \; = \; t\mathbf{x} + (1-t)\mathbf{z}$. Frequently we shall also say that the points $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ satisfy the *ordering relation* $\boldsymbol{\mathcal{B}}(\mathbf{x},\mathbf{y},\mathbf{z})$. The *closed segment* $[\mathbf{x};\mathbf{z}]$ consists of $\mathbf{x}$, $\mathbf{z}$, and all $\mathbf{y}$ such that $\mathbf{y}$ is between $\mathbf{x}$ and $\mathbf{z}$. In Exercise 1 below this is compared to the usual definition of closed interval in $\mathbb{R}$.

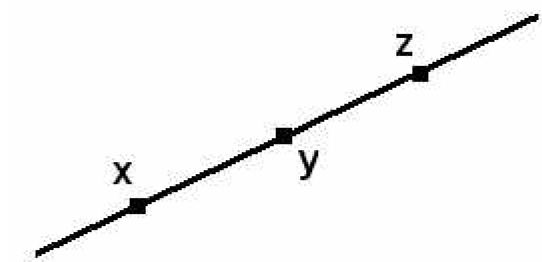


Figure VI.2

The *open segment* $(\mathbf{x};\mathbf{z})$ consists of all $\mathbf{y}$ such that $\mathbf{y}$ is between $\mathbf{x}$ and $\mathbf{z}$.

The next results show that our definition of betweenness satisfies some properties that are probably very apparent. However, since we are dealing with a fairly abstract setting, it is necessary to give rigorous proofs.

THEOREM VI.19. *Let $\mathbb{F}$ be an ordered field, let V be a vector space over $\mathbb{F}$, and let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ be distinct vectors in V. If $\boldsymbol{\mathcal{B}}(\mathbf{a},\mathbf{b},\mathbf{c})$ is true, then so is $\boldsymbol{\mathcal{B}}(\mathbf{c},\mathbf{b},\mathbf{a})$. However, each of the four statements $\boldsymbol{\mathcal{B}}(\mathbf{b},\mathbf{a},\mathbf{c})$, $\boldsymbol{\mathcal{B}}(\mathbf{c},\mathbf{a},\mathbf{b})$, $\boldsymbol{\mathcal{B}}(\mathbf{a},\mathbf{c},\mathbf{b})$, $\boldsymbol{\mathcal{B}}(\mathbf{b},\mathbf{c},\mathbf{a})$ is false.*

Proof. By assumption $\mathbf{b} \; = \; t\mathbf{a}+(1-t)\mathbf{c}$ for some t satisfying $0 < t < 1$. The latter inequalities imply $0 < (1-t) < 1$, and since $t \; = \; 1-(1-t)$ it follows that $\boldsymbol{\mathcal{B}}(\mathbf{c},\mathbf{b},\mathbf{a})$ is true.

The equation $\mathbf{b} \; = \; t\mathbf{a} + (1-t)\mathbf{c}$ (where $0 < t < 1$) implies that $-t\mathbf{a} \; = \; -\mathbf{b} + (1-t)\mathbf{c}$, which in turn means that

$$t^{-1}\mathbf{b} \; - \; t^{-1}(1-t)\mathbf{c} \; = \; t^{-1}\mathbf{b} \; + \; (1-t^{-1})\mathbf{c} \; .$$

Therefore $\mathbf{a} \; = \; s\mathbf{b} + (1-s)\mathbf{c}$ then implies $s = t^{-1}$. Since $0 < t < 1$ implies $t^{-1} > 1$, it follows that $\boldsymbol{\mathcal{B}}(\mathbf{b},\mathbf{c},\mathbf{a})$ is false. Furthermore, it also follows that $\boldsymbol{\mathcal{B}}(\mathbf{a},\mathbf{c},\mathbf{b})$ is false, for if the latter were true then by the preceding paragraph the order relation $\boldsymbol{\mathcal{B}}(\mathbf{b},\mathbf{c},\mathbf{a})$ would also be true.

Finally, $\mathbf{b} \; = \; t\mathbf{a} + (1-t)\mathbf{c}$ (where $0 < t < 1$) implies that $(t-1)\mathbf{c} \; = \; t\mathbf{a} - \mathbf{b}$, which in turn implies that

$$\mathbf{c} \; = \; \frac{t}{t-1}\mathbf{a} \; + \; \frac{-1}{t-1}\mathbf{b} \; .$$

Now $0 < t < 1$ implies $t-1 < 0$, so that

$$\frac{t}{t-1} \; < \; 0 \; .$$

The latter means that $\boldsymbol{\mathcal{B}}(\mathbf{c},\mathbf{a},\mathbf{b})$ is false, and as in the previous paragraph it follows that $\boldsymbol{\mathcal{B}}(\mathbf{a},\mathbf{c},\mathbf{b})$ is also false.■

THEOREM VI.20. *Let $\mathbb{F}$ and V be as above, and let* $\mathbf{a}$ *and* $\mathbf{b}$ *be distinct vectors in V. Then $\mathbf{c} \in V$ lies on the line* $\mathbf{ab}$ *if and only if one of $\mathbf{c} = \mathbf{a}$, $\mathbf{c} = \mathbf{b}$, $\boldsymbol{\mathcal{B}}(\mathbf{a},\mathbf{b},\mathbf{c})$, $\boldsymbol{\mathcal{B}}(\mathbf{c},\mathbf{a},\mathbf{b})$ or $\boldsymbol{\mathcal{B}}(\mathbf{b},\mathbf{c},\mathbf{a})$ is true. Furthermore, these conditions are mutually exclusive.*

Proof. We know that $\mathbf{c} \in \mathbf{ab}$ if and only if $\mathbf{c} = t\mathbf{a} + (1-t)\mathbf{b}$ for some t. We claim the five conditions are equivalent to $t = 1$, $t = 0$, $t < 0$, $t > 1$ and $0 < t < 1$ respectively. Thus it will suffice to verify the following:

(1) $\boldsymbol{\mathcal{B}}(\mathbf{a},\mathbf{b},\mathbf{c})$ is true if and only if $\mathbf{c} = t\mathbf{a} + (1-t)\mathbf{b}$ for some $t < 0$.

(2) $\boldsymbol{\mathcal{B}}(\mathbf{c},\mathbf{a},\mathbf{b})$ is true if and only if $\mathbf{c} = t\mathbf{a} + (1-t)\mathbf{b}$ for some $t > 1$.

PROOF OF (1). The condition $\mathbf{c} = t\mathbf{a} + (1-t)\mathbf{b}$ with $t < 0$ is equivalent to

$$\mathbf{b} \;=\; \frac{t}{t-1}\,\mathbf{a} \;+\; \frac{-1}{t-1}\,\mathbf{c}\ .$$

The conclusion in this case follows because the map sending t to $t/(t-1)$ is a 1–1 correspondence from the unbounded set $\{\, u \in \mathbb{F} \mid u < 0 \,\}$ to the bounded open interval $\{\, v \in \mathbb{F} \mid 0 < v < 1 \,\}$.□

PROOF OF (2). The condition $\mathbf{c} = t\mathbf{a} + (1-t)\mathbf{b}$ with $t > 1$ is equivalent to

$$\mathbf{a} \;=\; \frac{1}{t}\,\mathbf{c} \;+\; \left(1 - \frac{1}{t}\right)\mathbf{b}\ .$$

The conclusion in this case follows because the map sending t to $1/t$ is a 1–1 correspondence from the unbounded set $\{\, u \in \mathbb{F} \mid u > 1 \,\}$ to the bounded open interval $\{\, v \in \mathbb{F} \mid 0 < v < 1 \,\}$.■

Betweenness and cross ratios

Not surprisingly, there are important relationships between the concept of betweenness and the notion of cross ratio. Here is the most basic result.

THEOREM VI.21. *Let $\mathbb{F}$ be an ordered field, and let $\mathrm{J} : \mathbb{F}^n \to \mathbb{FP}^n$ be the usual projective extension mapping. Then three collinear points* $\mathbf{a}$, $\mathbf{b}$ *and* $\mathbf{c}$ *of V satisfy the order relation $\boldsymbol{\mathcal{B}}(\mathbf{a},\mathbf{c},\mathbf{b})$ if and only if*

$$\mathsf{XR}\big(\,\mathrm{J}(\mathbf{a}), \mathrm{J}(\mathbf{b}), \mathrm{J}(\mathbf{c}), L_\infty\,\big) \;<\; 0$$

where L_∞ is the ideal point of the line L containing $\mathbf{a}$, $\mathbf{b}$ *and* $\mathbf{c}$.

Proof. By Theorem V.17, if $\mathbf{c} = t\mathbf{a} + (1-t)\mathbf{b}$ then

$$\mathsf{XR}\big(\,\mathrm{J}(\mathbf{a}), \mathrm{J}(\mathbf{b}), \mathrm{J}(\mathbf{c}), L_\infty\,\big) \;=\; \frac{t-1}{t}\ .$$

This is negative if $0 < t < 1$ because $t - 1 < 0 < t$. We claim that the cross ratio is positive if either $t < 0$ or $t > 1$. If $t > 1$, then $t - 1 > 0$ and therefore the cross ratio is positive. Similarly, if $t < 0$, then $t - 1 < t < 0$ implies that the cross ratio is positive.■

Affine transformations obviously preserve betweenness (see Exercise 10 below). However, if $\boldsymbol{\mathcal{B}}(\mathbf{a},\mathbf{b},\mathbf{c})$ in $\mathbb{F}^n$ and T is a projective collineation of $\mathbb{FP}^n$ such that the images $\mathbf{a}'$, $\mathbf{b}'$, $\mathbf{c}'$ of

$\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ under T lie in (the image of) $\mathbb{F}^n$, then $\mathcal{B}(\mathbf{a}', \mathbf{b}', \mathbf{c}')$ is not necessarily true. Specific examples are given by projective collineations which interchange $\mathbf{a}$ and $\mathbf{b}$, and send $\mathbf{c}$ to itself.

If one wants some aspect of order and betweenness which IS preserved by projective collineations, it is natural to try something involving the cross ratio, and the preceding result may be viewed as motivation for the following definition and theorem:

Definition. Let $\mathbb{F}$ be an ordered field, let V be a vector space over $\mathbb{F}$, and let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ be collinear vectors in V. Then $\mathbf{a}$ and $\mathbf{b}$ *separate* $\mathbf{c}$ and $\mathbf{d}$ if one of $\{\mathbf{c}, \mathbf{d}\}$ is between $\mathbf{a}$ and $\mathbf{b}$ but the other is not. We shall write this as $\mathsf{sep}(\mathbf{a}, \mathbf{b} : \mathbf{c}, \mathbf{d})$. It is trivial to see that $\mathsf{sep}(\mathbf{a}, \mathbf{b} : \mathbf{c}, \mathbf{d})$ is equivalent to $\mathsf{sep}(\mathbf{a}, \mathbf{b} : \mathbf{d}, \mathbf{c})$ and $\mathsf{sep}(\mathbf{c}, \mathbf{d} : \mathbf{a}, \mathbf{b})$ (and one can also derive several other equivalent cross ratio statements from these).

There is a very simple and important characterization of separation in terms of cross ratios.

THEOREM VI.22. *Let $\mathbb{F}$ be an ordered field, let V be a vector space over $\mathbb{F}$, and let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ be collinear vectors in V. Then $\mathsf{sep}(\mathbf{a}, \mathbf{b} : \mathbf{c}, \mathbf{d})$ is true if and only if*

$$\big(\, \mathrm{J}(\mathbf{a}), \mathrm{J}(\mathbf{b}), \mathrm{J}(\mathbf{c}), \mathrm{J}(\mathbf{d}) \,\big) \;<\; 0 \;.$$

Proof. *Suppose that $\mathsf{sep}(\mathbf{a}, \mathbf{b} : \mathbf{c}, \mathbf{d})$ is true.* Without loss of generality, we may assume that $\mathcal{B}(\mathbf{a}, \mathbf{c}, \mathbf{b})$ is true but $\mathcal{B}(\mathbf{a}, \mathbf{d}, \mathbf{b})$ is false (either this holds or else the corresponding statement with $\mathbf{c}$ and $\mathbf{d}$ interchanged is true – in the latter case, reverse the roles of the two points). Under these conditions we have $\mathbf{c} \;=\; t\mathbf{a} + (1-t)\mathbf{b}$ where $0 < t < 1$ and $\mathbf{d} \;=\; s\mathbf{a} + (1-s)\mathbf{b}$ where $s < 0$ or $s > 1$. By Theorem V.17 we have

$$\mathsf{XR}\big(\, \mathrm{J}(\mathbf{a}), \mathrm{J}(\mathbf{b}), \mathrm{J}(\mathbf{c}), \mathrm{J}(\mathbf{d}) \,\big) \;=\; \frac{s(1-t)}{t(1-s)} \;.$$

The sign of this cross ratio equals the sign of $s/(1-s)$, and the latter is negative if either $s < 0$ or $s > 1$.$\square$

Suppose that the cross ratio is negative. We need to show that one of s and t satisfies $0 < u < 1$ and the other does not. To do this, we eliminate all the other possibilities.

CASE 1. Suppose we have $0 < s, t < 1$. Then all the factors of the numerator and denominator are positive.

CASE 2. Suppose neither satisfies $0 < u < 1$. Then the previous argument shows that one of s and $1-s$ is positive and likewise for t and $1-t$. Therefore the formula above implies that the cross ratio must be positive.$\blacksquare$

We thus make the general definition in $\mathbb{FP}^n$ that two points A and B **separate** two points C and D on AB if and only if $\mathsf{XR}(A, B, C, D) \;<\; 0$. If all four of these points are ordinary, then Theorem 22 provides a geometrical description of separation. The cases where one point is ideal can be described using the following two special cases:

(1) $\mathsf{sep}\big(\, \mathrm{J}(\mathbf{a}), \mathrm{J}(\mathbf{b}) : \infty, \mathrm{J}(\mathbf{c}) \,\big)$ and $\mathsf{sep}(\mathrm{J}(\mathbf{a}), \mathrm{J}(\mathbf{b}) : \mathrm{J}(\mathbf{c}), \infty)$ hold if and only if $\mathcal{B}(\mathbf{a}, \mathbf{c}, \mathbf{b})$ is true (see Theorems V.12 and V.17).

(2) $\mathsf{sep}\big(\, \mathrm{J}(\mathbf{a}), \infty : \mathrm{J}(\mathbf{c}), \mathrm{J}(\mathbf{d}) \,\big)$ and $\mathsf{sep}\big(\, \infty, \mathrm{J}(\mathbf{a}) : \mathrm{J}(\mathbf{c}), \mathrm{J}(\mathbf{d}) \,\big)$ hold if and only if $\mathcal{B}(\mathbf{c}, \mathbf{a}, \mathbf{d})$ is true because

$$\mathsf{XR}\big(\, \mathrm{J}(\mathbf{a}), \infty : \mathrm{J}(\mathbf{c}), \mathrm{J}(\mathbf{d}) \,\big) \;=\; \mathsf{XR}\big(\, \mathrm{J}(\mathbf{d}), \mathrm{J}(\mathbf{c}) : \infty, \mathrm{J}(\mathbf{a}) \,\big) \;.$$

The following observation is an immediate consequence of the definitions:

> *Let A, B, C, D be distinct collinear points in $\mathbb{FP}^n$, and let T be a projective collineation of $\mathbb{FP}^n$. Then $\mathsf{sep}(A, B : C, D)$ is true if and only if $\mathsf{sep}\big(T(A), T(B) : T(C), T(D)\big)$ is true.*■

A comprehensive visualization of separation for points on a real projective line may be given as follows:

As indicated in the picture below, there is a standard 1–1 correspondence (*stereographic projection*) between the points of $\mathbb{RP}^1$ and the points on the circle Γ in $\mathbb{R}^2$ which is tangent to the x-axis at the origin and whose center is $(0, \frac{1}{2})$. An ordinary point with standard affine coordinate u is sent to the intersection of Γ with the line joining $(u, 0)$ to $(0, 1)$, and the point at infinity is sent to $(0, 1)$. It is straightforward to check that this map σ defines a 1–1 correspondence from $\mathbb{RP}^1$ to Γ.[3]

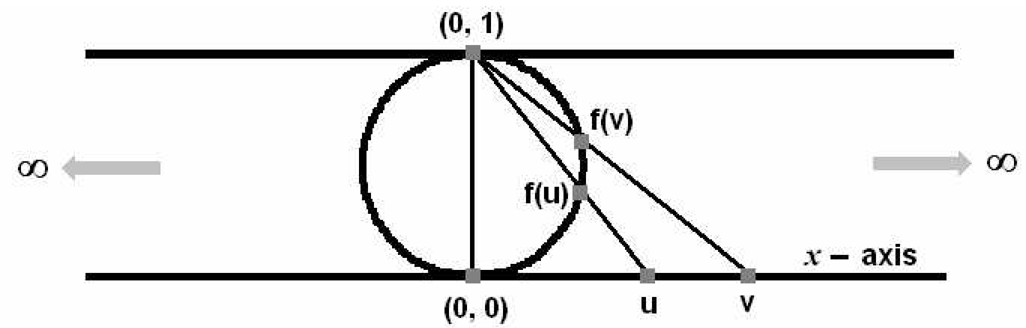


Figure VI.3

With respect to this correspondence, separation has the following interpretation. If $a,\ b \in \mathbb{RP}^1$, then $\Gamma - \{\sigma(a), \sigma(b)\}$ consists of two open arcs, and separation means that each arc contains exactly one of the points $\{c, d\}$.

We now summarize some basic properties of separation by means of the following theorem:

THEOREM VI.23. *If $\mathbb{F}$ is an ordered field and A, B, C, D are distinct collinear points of $\mathbb{FP}^n$, then the following hold:*

(*a*) $\mathsf{sep}(A, B : C, D)$ *implies* $\mathsf{sep}(A, B : D, C)$ *and* $\mathsf{sep}(C, D : A, B)$.

(*b*) *One and only one of the relations* $\mathsf{sep}(A, B : C, D)$, $\mathsf{sep}(B, C : D, A)$, *or and* $\mathsf{sep}(C, A : B, D)$ *is true.*

(*c*) *If* $\mathsf{sep}(A, B : C, D)$ *and* $\mathsf{sep}(B, C : D, E)$ *are true, then so is* $\mathsf{sep}(C, D : E, A)$.

(*d*) *If L is a line meeting AB, Y is a coplanar point on neither line, and X' is the intersection point of PX and L for $X = A, B, C, D$, then* $\mathsf{sep}(A, B : C, D)$ *implies* $\mathsf{sep}(A', B' : C', D')$.

[3]If we rotate the above picture about the y-axis in $\mathbb{R}^3$ we obtain a similar 1–1 correspondence between the complex projective line $\mathbb{CP}^1$ and the sphere of diameter 1 tangent to the xz-plane at the origin.

The proof is straightforward and is left as an exercise.■

One reason for listing the preceding four properties is that they come close to providing a complete characterization of separation.

THEOREM VI.24. *Let P be a Desarguian projective n-space, where $n \geq 2$, and suppose that P has an abstract notion of separation $\Sigma(\cdots,\cdots||\cdots,\cdots)$ which satisfies the four properties in the previous theorem. Assume that some (hence every) line contains at least four points. Then P is isomorphic to $\mathbb{FP}^n$, where $\mathbb{F}$ is an ordered skew-field, and the ordering of $\mathbb{F}$ has the property that* $\mathsf{sep}(A, B : C, D)$ *is true if and only if* $\Sigma(\cdots,\cdots||\cdots,\cdots)$ *is.*

In principle, this result is proved on pages 239–244 of Artzy, *Linear Geometry*. We say "in principle" because the result is only stated for projective planes in which Pappus' Theorem holds. However, the latter is not used explicitly in the argument on these pages,[4] and the restriction to planes is easily removed.

We would need only one more axiom to give a completely synthetic characterization of the real projective plane (and similarly for higher dimensional real projective spaces). Fairly readable formulations of the required *continuity condition* (as it is called) may be found in Coxeter, *The Real Projective Plane*, pages 161–162, and Artzy (*op. cit.*), page 244.

EXERCISES

Throughout these exercises $\mathbb{F}$ denotes an ordered field, and the ordering is given by the usual symbolism.

1. In the real numbers $\mathbb{R}$, prove that the closed interval $[a, b]$, consisting of all x such that $a \leq x \leq b$, is equal to the closed segment $[a; b]$ joining a to b as defined here, and likewise for their open analogs (a, b) and $(a; b)$. [*Hint:* If $a \leq c \leq b$ and $t = (b-a)/(c-a)$, consider $ta + (1-t)b$. If $c = ta + (1-t)b$ for $0 \leq t \leq 1$, why does this and $a \leq b$ imply that $a \leq c \leq b$?]

Definition. A subset $K \subset \mathbb{F}^n$ is **convex** if $\mathbf{x}$ and $\mathbf{y}$ in K imply that the closed segment $[\mathbf{x}; \mathbf{y}]$ is contained in K. — In physical terms for, say, $\mathbb{R}^2$ or $\mathbb{R}^3$, this means that K has "no dents or holes."

2. Prove that the following subsets of $\mathbb{F}$ are convex for an arbitrary $b \in \mathbb{F}$:

(i) The set $\{\, x \in \mathbb{F} \mid x > b\}$.

(ii) The set $\{\, x \in \mathbb{F} \mid x < b\}$.

(iii) The set $\{\, x \in \mathbb{F} \mid x \geq b\}$.

(iv) The set $\{\, x \in \mathbb{F} \mid x \leq b\}$.

3. Prove that the intersection of an arbitrary family of convex subsets of $\mathbb{F}^n$ is also convex.

4. Let $f : \mathbb{F}^n \to \mathbb{F}$ be a linear function of the form $f(\mathbf{x}) = \sum_i a_i x_i - b$.

[4] An explicit recognition that Pappus' Theorem is unnecessary appears in Forder, *Foundations of Euclidean Geometry*, pp. 196–197 and 203–206.

(a) Prove that for all $t \in \mathbb{F}$ we have $f\big(t\mathbf{x} + (1-t)\mathbf{y}\big) = t \cdot f(\mathbf{x}) + (1-t)f(\mathbf{y})$.

(b) Prove that if $K \subset \mathbb{F}^n$ is convex, then so is its image $f[K]$.

(c) Prove that if $C \subset \mathbb{F}$ is convex, then so is its inverse image $f^{-1}[C]$.

5. Let f be as in Exercise 4. Then the subsets of $\mathbb{F}^n$ on which f is positive and negative are called the *(two)* sides *of the hyperplane H_f* defined by f or the *(two)* half-spaces *determined by the hyperplane H_f*. Prove that each half-space is (nonempty and) convex, and if we have points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{F}^n$ such that $\mathbf{x}$ lies on one of the half-spaces and $\mathbf{y}$ lies on the other, then the closed segment $[\mathbf{x};\mathbf{y}]$ contains a point of the hyperplane H_f defined by the equation $f(\mathbf{z}) = 0$. — This statement is called the *hyperplane separation property* for $\mathbb{F}^n$.

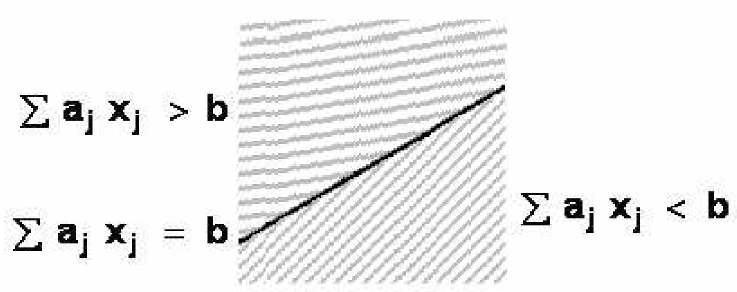


Figure VI.4

Also, explain why the hyperplane and its two sides are three pairwise disjoint subsets whose union is all of $\mathbb{F}^n$.

6. Formulate and prove a similar result to Exercise 5 for the set of all points in a k-plane $M \subset \mathbb{F}^n$ which are not in a $(k-1)$-plane $Q \subset M$.

7. Suppose that $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ are noncollinear points in $\mathbb{F}^2$. Define the **classical triangle $\Delta^c\mathbf{xyz}$** to be the union of the closed segments $[\mathbf{x};\mathbf{y}]$, $[\mathbf{x};\mathbf{z}]$, and $[\mathbf{y};\mathbf{z}]$. Prove the **Theorem of Pasch:**[5] A line L containing a point $\mathbf{w}$ in an open side $(\mathbf{x};\mathbf{y})$ of $\Delta^c\,\mathbf{xyz}$ either passes through $\mathbf{z}$ or else meets one of the other open sides $(\mathbf{x};\mathbf{z})$ or $(\mathbf{x};\mathbf{z})$. [*Hint:* Explain why $\mathbf{x}$ and $\mathbf{y}$ are on opposite sides of the L through $\mathbf{w}$. What can be said about $\mathbf{z}$ if it does not lie on this line?]

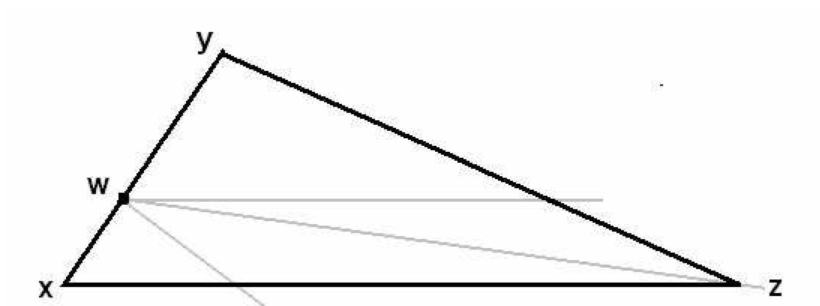


Figure VI.5

[5] Moritz Pasch (1843–1930) is mainly known for his work on the foundations of geometry, and especially for recognizing the logical deficiencies in Euclid's *Elements* and developing logically rigorous methods for addressing such issues. The theorem in the exercise is one example of a geometrical result that is tacitly assumed – but not proved — in the *Elements*.

7. If $\mathbf{x},\ \mathbf{y},\ \mathbf{z},\ \mathbf{w}$ are distinct points in $\mathbb{F}^2$ such that no three are noncollinear, then the *classical quadrilateral* $\Box^c\,\mathbf{xyzw}$ is the set

$$[\mathbf{x};\mathbf{y}]\ \cup\ [\mathbf{y};\mathbf{z}]\ \cup\ [\mathbf{z};\mathbf{w}]\ \cup\ [\mathbf{w};\mathbf{x}]\ .$$

It is called a *convex quadrilateral* if the following conditions hold:

$\mathbf{x}$ and $\mathbf{y}$ lie on the same side of $\mathbf{zw}$.
$\mathbf{y}$ and $\mathbf{z}$ lie on the same side of $\mathbf{wx}$.
$\mathbf{z}$ and $\mathbf{w}$ lie on the same side of $\mathbf{xy}$.
$\mathbf{w}$ and $\mathbf{x}$ lie on the same side of $\mathbf{yz}$.

The *diagonals* of a classical quadrilateral $\Box^c\,\mathbf{xyzw}$ are the segments $[\mathbf{x};\mathbf{z}]$ and $[\mathbf{y};\mathbf{w}]$. Prove that the diagonals of a convex quadrilateral have a point in common. Why must this point lie on $(\mathbf{x};\mathbf{z})\ \cap\ (\mathbf{y};\mathbf{w})$?

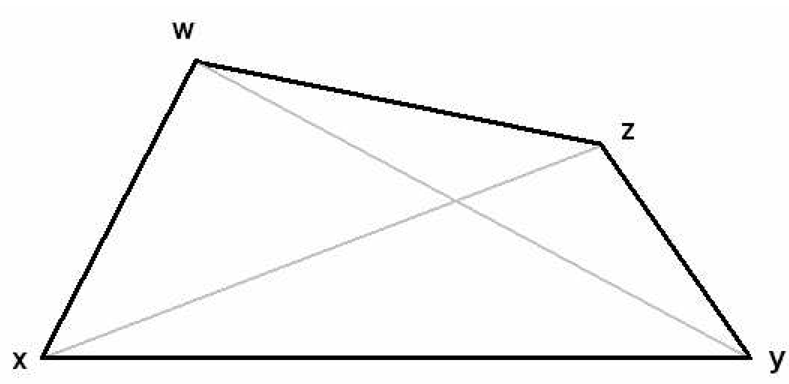


Figure VI.6

8. Give an explicit formula for the map defined by Figure VI.3 and the accompanying discussion.

9. Prove Theorem 24.

10. Suppose that $\mathbf{x},\ \mathbf{y},\ \mathbf{z}$ are points in $\mathbb{F}^n$ such that $\boldsymbol{\mathcal{B}}(\mathbf{x},\mathbf{y},\mathbf{z})$ is true and $T\in \mathrm{Aff}(\mathbb{F}^n)$. Prove that $\boldsymbol{\mathcal{B}}\big(\,T(\mathbf{x}),T(\mathbf{y}),T(\mathbf{z})\,\big)$ is also true.

11. In the notation of Exercise 5, let $\mathbf{y}_1,\ \cdots\ ,\mathbf{y}_n$ be an affine basis for the hyperplane H under consideration, and let $\mathbf{y}_0 \notin H$. Prove that $\mathbf{x}\in\mathbb{F}^n$ lies on the same side of H as $\mathbf{y}_0$ if the 0^{th} barycentric coordinate of $\mathbf{x}$ with respect to $\mathbf{y}_0,\ \mathbf{y}_1,\ \cdots\ ,\mathbf{y}_n$ (an affine basis for $\mathbb{F}^n$) is positive. What is the condition for $\mathbf{x}$ and $\mathbf{y}_0$ to lie on opposite sides?

CHAPTER VII

HYPERQUADRICS

Conic sections have played an important role in projective geometry almost since the beginning of the subject. In this chapter we shall begin by defining suitable projective versions of conics in the plane, quadrics in 3-space, and more generally hyperquadrics in n-space. We shall also discuss tangents to such figures from several different viewpoints, prove a geometric classification for conics similar to familiar classifications for ordinary conics and quadrics in $\mathbb{R}^2$ and $\mathbb{R}^3$, and we shall derive an enhanced duality principle for projective spaces and hyperquadrics. Finally, we shall use a mixture of synthetic and analytic methods to prove a famous classical theorem due to B. Pascal (1623–1662)[1] on hexagons inscribed in plane conics, a dual theorem due to C. Brianchon (1783–1864),[2] and several other closely related results.

1. Definitions

> The three familiar curves which we call the "conic sections" have a long history ... It seems that they will always hold a place in the curriculum. The beginner in analytic geometry will take up these curves after he has studied the circle. Whoever looks at a circle will continue to see an ellipse, unless his eye is on the axis of the curve. The earth will continue to follow a nearly elliptical orbit around the sun, projectiles will approximate parabolic orbits, [and] a shaded light will illuminate a hyperbolic arch. — **J. L. Coolidge (1873–1954)**

In classical Greek geometry, conic sections were first described synthetically as intersections of a plane and a cone. On the other hand, today such curves are usually viewed as sets of points (x, y) in the Cartesian plane which satisfy a nontrivial quadratic equation of the form

$$Ax^2 \;+\; 2Bxy \;+\; Cy^2 \;+\; 2D \;+\; 2E \;+\; F \;=\; 0$$

where at least one of $A,\ B,\ C$ is nonzero. In these notes we shall generally think of conics and quadrics in such terms. Here are some online references which relate the classical and modern approaches to these objects. The first contains some historical remarks, the second is a fairly detailed treatment which shows the equivalence of the classical and modern definitions only using material from elementary geometry, and the third contains a different proof that the definitions are equivalent using standard results from trigonometry.

`http://xahlee.org/SpecialPlaneCurves_dir/ConicSections_dir/conicSections.html`

`http://mathdl.maa.org/convergence/1/?pa=content&sa=viewDocument&nodeId=196&bodyId=60`

[1] Incidentally, he proved this result when he was 16 years old.

[2] This result was originally discovered without using duality.

`http://math.ucr.edu/~res/math153/history04Y.pdf`

The corresponding notion of *quadric surface* in $\mathbb{R}^3$ is generally defined to be the set of zeros of a nontrivial quadratic polynomial $p(x, y, z)$ in three variables (nontriviality means that at least one term of degree two has a nonzero coefficient). One can similarly define a *hyperquadric* in $\mathbb{R}^n$ to be the set of zeros of a nonzero quadratic polynomial $p(x_1, \cdots, x_n)$. Such an equation has the form

$$\sum_{i,j} a_{i,j} x_i x_j \ + \ 2 \cdot \sum_k b_k x_k \ + \ c \ = \ 0$$

where at least one of the coefficients $a_{i,j} \ = \ 0$.

One obvious question about our definitions is to give a concise but useful description of all the different types of conics, quadrics or hyperquadrics that exist in $\mathbb{R}^n$. Using linear algebra, in each dimension it is possible to separate or classify such objects into **finitely many** types such that

> *if Σ_1 and Σ_2 are hyperquadrics that are* affinely equivalent *(so that there is an affine transformation T of $\mathbb{R}^n$ such that $T[\Sigma_1] \ = \ \Sigma_2$, then Σ_1 and Σ_2 have the same type.* — In fact, *one can choose the affine transformation to have the form $T_1 \circ T_0$, where T_0 is a linear transformation and T_1 is given by a* diagonalizable *invertible linear transformation*; in other words, *there are nonzero scalars d_i such that for each i we have $T_1(\mathbf{e}_i) \ = \ d_i \mathbf{e}_i$, where $\mathbf{e}_i$ is the i^{th} standard unit vector in $\mathbb{R}^n$.*

For $n = 2$ and 3, the details of this classification are described explicitly in Section V.2 of the following online document:

`http://math.ucr.edu/~res/math132/linalgnotes.pdf`

The case of conics in $\mathbb{R}^2$ is summarized in the table on page 82 of this document, and the case of quadrics in $\mathbb{R}^3$ is summarized in the table on page 83 of the same document. In particular, there are fewer than 10 different types of possible nonempty figures in $\mathbb{R}^2$ (including degenerate cases of sets with one point or no points) and fewer than 20 different types of possible nonempty figures in $\mathbb{R}^3$ (also including an assortment degenerate cases). Later in this chapter we shall describe the analogous classification for $\mathbb{R}^n$ (with $n \geq 3$ arbitrary) in one of the exercises.

Projective extensions of hyperquadrics

We are now faced with an obvious question:

> *How does one define a hyperquadric in projective space?*

Let us consider the analogous situation in degree one. The sets of solutions to nontrivial linear equations $p(x_1, \cdots, x_n) = 0$ are merely hyperplanes. If $p(x_1, \cdots, x_n)s \ = \ \sum_i \ a_i x_i \ + \ b$, then this hyperplane is just the set of ordinary points in $\mathbb{RP}^n$ whose homogeneous coordinates satisfy the *homogeneous* linear equation

$$\sum_{i=1}^{n} a_i x_i \ + \ b x_{n+1} \ = \ 0 \ .$$

This suggests the following: Consider the quadratic polynomial

$$p(x_1, \cdots, x_n) \ = \ \sum_{i,j} a_{i,j} x_i x_j \ + \ 2 \cdot \sum_k b_k x_k \ + \ c$$

and turn it into a homogeneous quadratic polynomial by multiplying each degree 1 monomial in the summation by x_{n+1} and multiplying the constant term by x_{n+1}^2. We then obtain the modified quadratic polynomial

$$\overline{p}(x_1, \cdots, x_n) \;=\; \sum_{i,j} a_{i,j} x_i x_j \;+\; 2 \cdot \sum_{k} b_k x_k x_{n+1} \;+\; c x_{n+1}^2$$

which is homogeneous and has the following compatibility properties:

THEOREM VII.1. *(i) If X is a point in $\mathbb{RP}^n$ and ξ and ξ' are homogeneous coordinates for X, then $\overline{p}(\xi) = 0$ if and only if $\overline{p}(\xi') = 0$.*

(ii) The set of zeros for p is equal to the set of ordinary points in $\mathbb{RP}^n$ whose homogeneous coordinates are zeros of $\overline{p}$.

Proof. We shall proof the two parts separately.

PROOF OF (i). Observe that $\overline{p}(k\xi) \;=\; k^2 \cdot \overline{p}(\xi)$ by direct computation. Therefore $\xi' = k\xi$ for some $k \neq 0$ implies that $\overline{p}(\xi') \;=\; 0$ if and only if $\overline{p}(\xi) \;=\; 0.\square$

PROOF OF (ii). If $\mathbf{x} \in \mathbb{R}^{n,1}$, then the transpose of $(x_1, \cdots, x_n, 1)$ is a set of homogeneous coordinates for $\mathrm{J}(\mathbf{x}) \in \mathbb{RP}^n$, and it is elementary to check that the solutions to the equation $p = 0$ contained in the intersection of the set of ordinary points and the points in $\mathbb{RP}^n$ whose homogeneous coordinates are solutions to the equation $\overline{p} = 0$ (in particular, we have $p(x_1, \cdots, x_n) \;=\; \overline{p}(x_1, \cdots, x_n, 1))$. Conversely, if $\overline{p}(x_1, \cdots, x_n, x_{n+1}) = 0$ where $x_{n+1} \neq 0$, then we also have

$$0 \;=\; \frac{1}{x_{n+1}^2} \cdot \overline{p}(x_1, \cdots, x_n, x_{n+1}) \;=\; \overline{p}\left(\frac{x_1}{x_{n+1}}, \cdots, \frac{x_n}{x_{n+1}}, 1\right) \;=\; p(x_1, \cdots, x_n)$$

and hence the solutions to $\overline{p} = 0$ in the image of J are all ordinary points which are solutions to $p = 0.\blacksquare$

All of the preceding discussion makes at least formal sense over an arbitrary field $\mathbb{F}$; of course, the mathematical value of the quadrics considered depends strongly upon the solvability of quadratic equations within the given field.[3] Define a *hyperquadric* Σ in $\mathbb{FP}^n$ to be the set of zeros of a homogeneous quadratic equation:

$$\sum_{i,j=1}^{n+1} a_{i,j} x_i x_j \;=\; 0$$

In the study of hyperquadrics we generally assume that $1 + 1 \neq 0$ in $\mathbb{F}$. This condition allows us to choose the n^2 coefficients $a_{i,j}$ so that $a_{i,j} = a_{j,i}$; for if we are given an arbitrary homogeneous quadratic equation as above and set $b_{i,j} \;=\; \frac{1}{2}(a_{j,i} + a_{i,j})$, then it is easy to see that

$$\sum_{i,j=1}^{n+1} a_{i,j} x_i x_j \;=\; 0 \text{ if and only if } \sum_{i,j=1}^{n+1} b_{i,j} x_i x_j \;=\; 0$$

because we have

$$\sum_{i,j=1}^{n+1} b_{i,j} x_i x_j \;=\; \tfrac{1}{2}\left(\sum_{i,j=1}^{n+1} a_{i,j} x_i x_j \;+\; \sum_{i,j=1}^{n+1} a_{j,i} x_i x_j\right) .$$

[3] All fields in this chapter are assumed to have commutative multiplications.

For these reasons, *we shall henceforth assume* $1 + 1 \neq 0$ *in* $\mathbb{F}$ *and* $a_{i,j} = a_{j,i}$ *for all* i *and* j.

It is natural to view the coefficients $a_{i,j}$ as the entries of a *symmetric* $(n+1) \times (n+1)$ matrix A. If we do so and Σ is the hyperquadric in $\mathbb{FP}^n$ defined by the equation $\sum_{i,j} a_{i,j} x_i x_j = 0$, then we may rewrite the defining equation for Σ as follows: A point X lies on Σ if and only if for some (equivalently, for all) homogeneous coordinates ξ representing X we have

$$ {}^{\mathbf{T}}\xi A \xi \ = \ 0 \ . $$

If we have an affine quadric in $\mathbb{F}^n$ defined by a polynomial p as above, then an $(n+1) \times (n+1)$ matrix defining its projective extension is given in block form by

$$ \begin{pmatrix} A & {}^{\mathbf{T}}\mathbf{b} \\ \mathbf{b} & c \end{pmatrix} $$

where the symmetric matrix $A = (\, a_{i,j} \,)$ gives the second degree terms of p, the row vector $2 \cdot \mathbf{b}$ gives the first degree terms b_i (note the coefficient!), and c gives the constant term.

Hypersurfaces of higher degree

The reader should be able to define projective hypercubics, hyperquartics, *etc.*, as well as the projective hyper—ic associated to an affine hyper—ic. Subsets of these types are generally called **projctive algebraic varieties**; they have been studied extensively over the past 300 years and have many interesting and important properties. The mathematical study of such objects has remained an important topic in mathematics ever since the development of projective geometry during the 19$^{\text{th}}$ century, but it very quickly gets into issues far beyond the scope of these notes. In particular, the theory involves a very substantial amount of input from multivariable calculus and the usual approaches also require considerably more sophisticated algebraic machinery than we introduce in these notes. The rudiments of the theory appear in Sections V.4–V.6 of the book by Bumcrot, and a more complete treatment at an advanced undergraduate level is given in Seidenberg, *Elements of the Theory of Algebraic Curves*, as well as numerous other introductory books on **algebraic geometry.**

Projective algebraic varieties also turn out to have important applications in various directions, including issues in theoretical physics, the theory of encryption, and even the proof of Fermat's Last Theorem during the 1990s which was mainly due to Andrew Wiles (the word "mainly" is included because the first complete proof required some joint work of Wiles with R. Taylor, and Wiles' work starts with some important earlier results by others). A reader who wishes to learn more about some of these matters may do so by going to the final part of Section IV.5 in the online document

`http//:math.ucr.edu~res/math133/coursenotes4b.pdf`

and checking the traditional and electronic references cited there.

EXERCISES

1. Consider the conics in $\mathbb{R}^2$ defined by the following equations:

(i) The circle defined by $x^2 + y^2 - 1 = 0$.

(ii) The hyperbola defined by $xy - 1 = 2$.

(iii) The parabola defined by $y - x^2 = 0$.

Show that the associated projective conics have 0, 1 and 2 points at infinity respectively, and give homogeneous coordinates for these points.

2. Find which points (if any) at infinity belong to the projective conics associated to the conics in $\mathbb{R}^2$ defined by the following equations.

(i) $x^2 - 2y^2 - 2xy \ = \ 0$

(ii) $3x^2 + 4y^2 - 4x + 2 \ = \ 0$

(iii) $x^2 + y^2 - 4y \ = \ 4$

(iv) $x^2 - 4xy - 4y^2 - 2y \ = \ 4$

3. Find the points at infinity on the projective quadrics associated to the quadrics in $\mathbb{R}^3$ defined by the following equations.

(i) $x^2 + y^2 - z^2 \ = \ 1$

(ii) $x^2 + y^2 - z^2 - 6x - 8y \ = \ 0$

(iii) $x^2 + y^2 \ = \ 2z$

(iv) $x^2 - y^2 - z^2 \ = \ 1$

(v) $x^2 + y^2 \ = \ z$

(vi) $x^2 + y^2 \ = \ z^2$

4. For each of the following affine quadrics σ in $\mathbb{R}^3$, find a symmetric 4×4 matrix such that the projective extension $\mathbb{P}(\Sigma)$ of Σ is defined by the equation ${}^{\mathbf{T}}\xi A \xi \ = \ 0$.

(i) Σ is defined by the affine equation $4x^2 + 3y^2 - z^2 + 2x + y + 2z - 1 \ = \ 0$.

(ii) Σ is defined by the affine equation $3x^2 + y^2 + 2z^2 + 3x + 3y + 4z \ = \ 0$.

(iii) Σ is defined by the affine equation $2x^2 + 4z^2 - 4x - y - 24z + 36 \ = \ 0$.

(iv) Σ is defined by the affine equation $4x^2 + 9y^2 + 5z^2 - 4xy + 8yz + 12xz + 9z - 3 \ = \ 0$.

2. Tangents

Tangent lines to circles play an important role in classical Euclidean geometry, and their generalizations to other conics we also known to classical Greek mathematicians such as Archimedes (287 B. C. E. – 212 B. C. E.) and Apollonius of Perga (*c.* 262 B. C. E. – *c.* 190 B. C. E.). In modern mathematics they are generally defined using concepts and results from single variable or multivariable differential calculus. Of course, the latter is designed to work primarily in situations where the coordinates are real or complex numbers, and since we want to consider more general coordinates we need to develop an approach that is at least somewhat closer to the classical viewpoint.

In these notes we shall concentrate on the following two ways of viewing tangents to conics in $\mathbb{R}^2$ or quadrics in $\mathbb{R}^3$.

1. SYNTHETIC APPROACH. A line is tangent to a hyperquadric if and only if it lies wholly in the hyperquadric or has precisely one point of intersection with the hyperquadric.
2. ANALYTIC APPROACH. Let $X \in \Sigma \cap L$, where Σ is a hyperquadric and L is a line. Then L is tangent to Σ if and only if there is a differentiable curve $\gamma : (a;b) \to \mathbb{R}^n$ lying totally in Σ such that $\gamma(t_0) = \mathbf{x}$ for some $t_0 \in (a;b)$ and L is the line $\mathbf{x} \ + \ \mathbb{R} \cdot \gamma'(t_0)$.

For our purposes the first viewpoint will be more convenient; in Appendix E we shall show that the analytic approach is consistent with the synthetic viewpoint, at least in all the most important cases. Actually, the viewpoint of calculus is the better one for generalizing tangents to cubics, quartics, *etc.*, but a correct formulation is too complicated to be given in these notes.

We begin with a result on solutions to homogeneous quadratic equations in two variables:

THEOREM VII.2. *Suppose that $\mathbb{F}$ is a field in which $1+1 \neq 0$, and (x_1, y_1), (x_2, y_2), (x_3, y_3) are solutions to the homogeneous quadratic equation*

$$ax^2 \ + \ bxy \ + \ cy^2 \ = \ 0 \ .$$

Then either $a = b = c = 0$ or else one of (x_1, y_1), (x_2, y_2), (x_3, y_3) is a nonzero multiple of another.

Proof. If the hypothesis holds, then in matrix terminology we have

$$\begin{pmatrix} x_1^2 & x_1y_1 & y_1^2 \\ x_2^2 & x_2y_2 & y_2^2 \\ x_3^2 & x_3y_3 & y_3^2 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} \ = \ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \ .$$

Suppose not all of $a, \ b, \ c$ are nonzero. Then the given 3×3 matrix is not invertible and hence has a zero determinant. But the determinant of such a matrix may be computed directly, and up to a sign factor it is equal to

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \cdot \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} \cdot \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix}$$

The vanishing of this determinant implies that one of the 2×2 determinants in the factorization must be zero, and the latter implies that the rows of the associated 2×2 matrix are proportional to each other.■

The preceding result has the following important geometric application:

THEOREM VII.3. *Let Σ be a hyperquadric in $\mathbb{RP}^n$, let $X \in \Sigma$, and let L be a line containing X. Then $\Sigma \cap L$ is either $\{X\}$, two points, or all of L.*

Proof. Let $Y \neq X$ where $Y \in L$, let ξ and η denote homogeneous coordinates for X and Y respectively, and suppose that Σ is defined by the equation

$$ {}^{\mathbf{T}}\omega A \omega \;=\; 0 $$

where A is a symmetric $(n+1) \times (n+1)$ matrix and ω represents $W \in \mathbb{RP}^n$.

If $Z \in L$ and is represented by the homogeneous coordinates ζ, then $\zeta \;=\; u\xi + v\eta$ for some $u, \; v \in \mathbb{F}$ that are not both zero. By construction, $Z \in \Sigma$ if and only if

$$ 0 \;=\; {}^{\mathbf{T}}\zeta A \zeta \;=\; {}^{\mathbf{T}}(u\xi + v\eta) A (u\xi + v\eta) \;=\; $$

$$ u^2\,{}^{\mathbf{T}}\xi A \xi \;+\; 2uv\,{}^{\mathbf{T}}\eta A \xi \;+\; v^2\,{}^{\mathbf{T}}\eta A \eta \;=\; u^2 p \;+\; 2uvq \;+\; v^2 q $$

for suitable constants $p, \; q, \; r$. We claim that $\Sigma \cap L$ has at least three points if and only if $L \subset \Sigma$. The "only if" implication is trivial, so we shall focus on the "if" direction. — Suppose that $Z_1, \; Z_2, \; Z_3$ are points on $\Sigma \cap L$, and take homogeneous coordinates $\zeta_i \;=\; u_i\xi + v_i\eta$ for Z_i. By Theorem 2, either $p = q = r = 0$ (in which case $L \subset \Sigma$) or else one of the pairs (u_i, v_i) is proportional to the other, say $(u_j, v_j) = m(u_k, v_k)$ for some $m \neq 0$. In this case we have that $Z_j = Z_k$ and hence $Z_1, \; Z_2, \; Z_3$ are not distinct.■

Definition. Let Σ be a hyperquadric, let $X \in \Sigma$, and let L be a line containing X. We shall say that L is a *tangent line* to Σ at X if either $\Sigma \cap L - \{X\}$ or $L \subset \Sigma$. In the remaining case where $\Sigma \cap L$ consists of two points, we shall say that L is a *secant line through* X. The *tangent space* to Σ at X is equal to the union of all tangent lines to Σ at X.

Singular and nonsingular points

If we consider the conic in $\mathbb{R}^2$ defined by the eqution $x^2 - y^2 = 0$ we see that the structure of the conic at the origin is different than at other points, for the conic is given by a pair of lines which intersect at the origin. Some words which may be used to describe this difference are exceptional, special or *singular*. A concise but informative overview of singular points for plane curves appears in the following online reference:

`http://mathworld.wolfram.com/SingularPoint.html`

There are corresponding theories of singularities for surfaces in $\mathbb{R}^3$, and more generally for hypersurfaces in $\mathbb{R}^n$. Not surprisingly, if one is only interested in hyperquadrics as in these notes, then everything simplifies considerably. We shall explain the relationship between the theory of singular and nonsingular points for hyperquadrics and the general case in Appendix E.

We have given a purely synthetic definition of the tangent space to a hyperquadric $\Sigma \subset \mathbb{FP}^n$ at a point $X \in \Sigma$. The first step is to give an algebraic description of the tangent space in terms of homogeneous coordinates.

THEOREM VII.4. *Let $\mathbb{F}$ and $\Sigma \subset \mathbb{FP}^n$ be as above, and let $X \in \Sigma$. Then the tangent space to Σ at X is either a hyperplane in $\mathbb{FP}^n$ or all of $\mathbb{FP}^n$. In the former case, X is said to be a* nonsingular *point, and in the latter case X is said to be a* singular point. *Furthermore, if Σ is defined by the symmetric matrix A and ξ is a set of homogeneous coordinates for X, then in the nonsingular case ${}^{\mathbf{T}}\xi A$ is a (nonzero) set of homogeneous coordinates for the tangent hyperplane, but in the singular case we have ${}^{\mathbf{T}}\xi A = \mathbf{0}$.*

EXAMPLES. Suppose we consider the projectivizations of the circle $x^2 + y^2 = 1$, the hyperbola $x^2 - y^2 = 1$, the parabola $y = x^2$, and the pair of intersecting lines $x^2 = y^2$. Then the corresponding projective conics are defined by the following homogeneous quadratic equations:

$$x_1^2 + x_2^2 - x_3^2 = 0, \qquad x_1^2 - x_2^2 - x_3^2 = 0$$

$$x_1^2 - x_2 x_3 = 0, \qquad x_1^2 - x_2^2 = 0$$

In the first three cases the associated 3×3 symmetric matrix A is invertible, and hence ${}^{\mathbf{T}}\xi A \neq \mathbf{0}$ for all nonzero ξ, so that every point of these projective conics will be a nonsingular point. — On the other hand, in the fourth example, the symmetric matrix A is not invertible, and in fact its kernel (either on the left or right side!) consists of all vectors whose first and second coordinates are equal to zero. This implies that all points on the conic except $\mathrm{J}(\mathbf{0})$ are nonsingular but $\mathrm{J}(\mathbf{0})$ is singular. These examples are all consistent with our intuition that the first three curves behave regularly (or are nonsingular) at all points and the fourth curve behaves regularly at all points except the origin.

Proof. In the proof of the preceding theorem, we noted that if $Y \in \mathbb{FP}^n$ with homogeneous coordinates η and $Z \in XY$ has homogeneous coordinates $\zeta = u\xi + v\eta$, then $Z \in \Sigma$ if and only if

$$u^2 \left({}^{\mathbf{T}}\xi A \xi\right) + 2uv \left({}^{\mathbf{T}}\xi A \eta\right) + v^2 \left({}^{\mathbf{T}}\eta A \eta\right) = 0$$

and the number of points on $XY \cap \Sigma$ depends upon the equivalence classes of solutions to this equation, which we shall call the INTERSECTION EQUATION.

CLAIM: The line XY is tangent to Σ if and only if ${}^{\mathbf{T}}\xi A \eta = {}^{\mathbf{T}}\eta A \xi = 0$.

Suppose first that XY is tangent to Σ. If XY is contained in Σ, then we have

$${}^{\mathbf{T}}\xi A \xi = {}^{\mathbf{T}}\eta A \eta = {}^{\mathbf{T}}(\xi + \eta) A (\xi + \eta) = 0$$

and elementary manipulations of these equations show that $2 \cdot {}^{\mathbf{T}}\eta A \xi = 0$. On the other hand, if $XY \cap \Sigma = \{X\}$, then ${}^{\mathbf{T}}\eta A \eta = 0$ and the only solutions to the Intersection Equation in the first paragraph of the proof are pairs (u, v) which are nonzero scalar multiples of $(1, 0)$. Therefore, the Intersection Equation evaluated at $(1, t)$ is equal to zero if and only if $t = 0$. However, it is easy to check that the ordered pair

$$\left(1, \; -\frac{{}^{\mathbf{T}}\xi A \eta}{{}^{\mathbf{T}}\eta A \eta}\right)$$

solves the Intersection Equation because

$${}^{\mathbf{T}}\xi A \xi = 0$$

and therefore we must have ${}^{\mathbf{T}}\xi A \eta = {}^{\mathbf{T}}\eta A \xi = 0$.□

Conversely, suppose that ${}^{\mathbf{T}}\xi A \eta = {}^{\mathbf{T}}\eta A \xi = 0$. Since ${}^{\mathbf{T}}\xi A \xi = 0$, the Intersection Equation reduces to

$$v^2 \left({}^{\mathbf{T}}\eta A \eta\right) = 0 \; .$$

This equation means that either ${}^{\mathbf{T}}\eta A\eta \;=\; 0$, in which case we have $L \subset \Sigma$, or else $v = 0$, in which case every solution (u,v) of the Intersection equation is proportional to the known solution $(1,0)$, so that $\Sigma \cap XY \;=\; \{X\}$.

To conclude the proof, we have shown that the tangent space at X is the set of all points Y such that ${}^{\mathbf{T}}\xi A\eta \;=\; 0$. If ${}^{\mathbf{T}}\xi A \;=\; \mathbf{0}$, this is all of $\mathbb{FP}^n$, and if ${}^{\mathbf{T}}\xi A \;\neq\; \mathbf{0}$, this is the hyperplane with homogeneous coordinates ${}^{\mathbf{T}}\xi A$.■

We shall say that a hyperquadric Σ is **nonsingular** if for each $X \in \Sigma$ the tangent space at X is a hyperplane (algebraically, this means that if ξ represents X then ${}^{\mathbf{T}}\xi A \neq \mathbf{0}$.

THEOREM VII.5. *If Σ is a hyperquadric defined by the symmetric matrix A, then Σ is nonsingular if and only if A is invertible.*

Proof. Suppose first that A is invertible. Then $\xi \neq \mathbf{0}$ implies that ${}^{\mathbf{T}}\xi A$ is nonzero, and by the preceding result it follows that the tangent space at every point must be a hyperplane.□

Conversely, suppose that A is not invertible. Then there is some $\xi \neq \mathbf{0}$ such that ${}^{\mathbf{T}}\xi A \;=\; \mathbf{0}$, and if ξ represents X it follows that $X \in \Sigma$ and X is a singular point of Σ.■

By definition, each symmetric matrix A determines a hyperquadric Σ_A. This is not a 1–1 correspondence, for if c is a nonzero scalar then clearly $\Sigma_A \;=\; \Sigma_{cA}$. We shall now use the notion of tangent hyperplane to show that, in many cases, this is the only condition under which two matrices can define the same hyperquadric. Further discussion of this question is given in Section 2 of Appendix E.

THEOREM VII.6. *Let A and B be symmetric $(n+1)\times(n+1)$ matrices over the field $\mathbb{F}$ in which $1+1 \neq 0$, and suppose they define the same nonempty hyperquadric in $\mathbb{FP}^n$. Assume that Σ has at least one nonsingular point. Then B is a scalar multiple of A.*

Proof. We are given that Σ has a nonsingular point X; let ξ be a set of homogeneous coordinates for X. Then both ${}^{\mathbf{T}}\xi A$ and ${}^{\mathbf{T}}\xi B$ define the same hyperplane and hence ${}^{\mathbf{T}}\xi A \;=\; k\cdot{}^{\mathbf{T}}\xi B$ for some nonzero scalar k.

Suppose now that Y does not lie on this tangent hyperplane, and let η be a set of homogeneous coordinates for Y. Then the line XY meets Σ in a second point which has homogeneous coordinates of the form $u\xi + \eta$ for some $u \in \mathbb{F}$. This scalar satisfies the following equations:

$$2u\,{}^{\mathbf{T}}\xi A\eta \;+\; {}^{\mathbf{T}}\eta A\eta \;=\; 0, \qquad\qquad 2u\,{}^{\mathbf{T}}\xi B\eta \;+\; {}^{\mathbf{T}}\eta B\eta \;=\; 0$$

Since ${}^{\mathbf{T}}\xi A \;=\; k\cdot{}^{\mathbf{T}}\xi B$ the equations above imply that

$${}^{\mathbf{T}}\eta A\eta \;=\; k\cdot{}^{\mathbf{T}}\eta B\eta$$

for all Y whose homogeneous coordinates satisfy ${}^{\mathbf{T}}\xi A\eta \neq 0$ (*i.e.*, all vectors in $\mathbb{F}^{n+1,1}$ except those in the n-dimensional subspace defined by the tangent hyperplane to Σ and X).

To prove that ${}^{\mathbf{T}}\eta A\eta \;=\; k\cdot{}^{\mathbf{T}}\eta B\eta$ if Y lies in the tangent hyperplane at X, let Z be a point which is not on the tangent hyperplane. Then

$${}^{\mathbf{T}}\omega A\omega \;=\; k\cdot{}^{\mathbf{T}}\omega B\omega$$

for $\omega = \zeta,\ \eta+\zeta,\ \eta-\zeta$. Let $C = A$ or B, and write $\Psi_C(\gamma,\delta) = {}^{\mathbf{T}}\gamma C\delta$. We then have the following:

$$\psi_C(\eta,\zeta) = \tfrac{1}{4}\Psi_C(\eta+\zeta,\eta+\zeta) - \tfrac{1}{4}\Psi_C(\eta-\zeta,\eta-\zeta)$$

$$\Psi_C(\eta,\eta) = \Psi_C\big(\,(\eta+\zeta)-\zeta,\ (\eta+\zeta)-\zeta\,\big)$$

By the first of these and the preceding paragraph, we have $\Psi_A(\eta,\zeta) = k\cdot\Psi_B(\eta,\zeta)$. Using this, the second equation above and the preceding paragraph, we see that $\Psi_A(\eta,\eta) = k\cdot\Psi_B(\eta,\eta)$ if η represents a point Y in the tangent hyperplane to Σ at X. Applying this and the first displayed equation to arbitrary nonzero vectors $\eta,\ \zeta\in\mathbb{F}^{n+1,1}$, we see that $\Psi_A(\eta,\zeta) = k\cdot\Psi_B(\eta,\zeta)$. Since $c_{i,j}$ is the value of $\Psi_C(\mathbf{e}_i,\mathbf{e}_j)$ if $\mathbf{e}_i$ and $\mathbf{e}_j$ are the standard unit vectors (the k^{th} coordinate of $\mathbf{e}_k$ is 1 and the rest are 0), we see that $a_{i,j} = k\cdot b_{i,j}$ for all i and j, and hence we see that $B = k\cdot A$.■

EXERCISES

In all these exercises $\mathbb{F}$ denotes a (commutative) field in which $1+1\neq 0$.

1. Find the singular points (if any) of the projective conics given in Exercise 3 of the previous section.

2. Find the equations of the tangent lines tot he following conics in $\mathbb{RP}^2$ at the indicated points:

(i) The conic defined by $x_1^2 + 2x_1x_2 + 4x_1x_3 + 3x_2^2 - 12x_1x_3 + 2x_3^2 = 0$ at the points

$$\begin{pmatrix}1\\1\\1\end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix}1\\1\\3\end{pmatrix}.$$

(ii) The conic defined by $x_1^2 - 2x_1x_2 + 4x_2^2 - 4x_3^2 = 0$ at the points

$$\begin{pmatrix}2\\2\\-1\end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix}2\\0\\1\end{pmatrix}.$$

Definition. Let Σ be a hyperquadric in $\mathbb{FP}^n$ defined by the $(n+1)\times(n+1)$ matrix A such that Σ has at least one nonsingular point. Two points X and Y in $\mathbb{FP}^n$ are said to be *conjugate* with respect to Σ if they have homogeneous coordinates ξ and η respectively such that ${}^{\mathbf{T}}\xi A\eta = 0$. By Theorem 6, this definition does not depend upon any of the choices (including A). Moreover, a point is self-conjugate if and only if it lies on Σ.

3. In the setting above, assume that $X\notin\Sigma$ and Y is conjugate to P with respect to Σ. Suppose that $XY\cap\Sigma$ consists of two points, say A and B. Prove that $\mathsf{XR}(X,Y,A,B) = -1$.

Note. If Σ is nonsingular and nonempty (hence A is invertible by Theorem 5) and $X \in \mathbb{FP}^n$, then a hyperplan with homogeneous coordinates ${}^{\mathbf{T}}\xi A$ is called the *polar hyperplane* of X with respect to Σ. The map $\mathbf{P}$ sending X to its polar hyperplane is a collineation from $\mathbb{FP}^n$ to its dual $\left(\mathbb{FP}^n\right)^*$ is called a **polarity**, and it has the property that the composite

$$\mathbb{FP}^n \xrightarrow[\cong]{\mathbf{P}} (\mathbb{FP}^n)^* \xrightarrow[\cong]{\mathbf{P}^*} (\mathbb{FP}^n)^{**}$$

is the identity.

4. Let Σ be an affine hyperquadric in $\mathbb{F}^n$, where $n \geq 3$, and suppose that L is a line in $\mathbb{F}^n$ such that $L \subset \Sigma$. Denote the projective extension of Σ by Σ^*. Prove that the ideal point L_∞, and in fact the entire projective line

$$\mathrm{J}[L] \;\cup\; \{L_\infty\}$$

is contained in Σ^*. [*Hint:* The field $\mathbb{F}$ contains at least three elements. What does this imply about the number of points on L, and how does this lead to the desired conclusion?]

3. Bilinear forms

At this point it is convenient to discuss a topic in linear algebra which is generally not covered in first courses on the subject. For the time being, $\mathbb{F}$ will be a (commutative field with no assumption on whether or not $1+1=0$ or $1+1\neq 0$.

Definition. Let V be a vector space over $\mathbb{F}$. A *bilinear form* on $\mathbb{F}$ is a function

$$\Phi : V \times V \longrightarrow \mathbb{F}$$

with the following properties:

(Bi–1) $\Phi(\mathbf{v}+\mathbf{v}',\mathbf{w}) = \Phi(\mathbf{v},\mathbf{w}) + \Phi(\mathbf{v}',\mathbf{w})$ for all $\mathbf{v},\ \mathbf{v}',\ \mathbf{w} \in V$.

(Bi–2) $\Phi(\mathbf{v},\mathbf{w}+\mathbf{w}') = \Phi(\mathbf{v},\mathbf{w}) + \Phi(\mathbf{v},\mathbf{w}')$ for all $\mathbf{v},\ \mathbf{w},\ \mathbf{w}' \in V$.

(Bi–3) $\Phi(c\cdot\mathbf{v},\mathbf{w}) = c\cdot\Phi(\mathbf{v},\mathbf{w}) = \Phi(\mathbf{v},c\cdot\mathbf{w})$ for all $\mathbf{v},\ ,\mathbf{w} \in V$ and $c \in \mathbb{F}$.

The reader will notice the similarities between the identities for Φ and the identities defining the dot product on $\mathbb{R}^n$. Both are scalar valued, distributive in both variables, and homogeneous (of degree 1) with respect to scalars. However, we are not assuming that Φ is commutative — in other words, we make no assumption about the difference between $\Phi(\mathbf{v},\mathbf{w})$ and $\Phi(\mathbf{w},\mathbf{v})$ — and we can have $\Phi(\mathbf{x},\mathbf{x}) = 0$ even if $\mathbf{x}$ is nonzero.

EXAMPLES. **1.** Let $\mathbb{F}=\mathbb{R}$ and $V=\mathbb{R}^2$, and let $\Phi(\mathbf{x},\mathbf{y}) = x_1y_2 - x_2y_1$, where by convention $\mathbf{a}\in\mathbb{R}^2$ can be written in coordinate form as (a_1,a_2). Then $\Phi(\mathbf{y},\mathbf{x}) = -\Phi(\mathbf{x},\mathbf{y})$ for all $\mathbf{x}$ and $\mathbf{y}$ and we also have $\Phi(\mathbf{z},\mathbf{z}) = 0$ for all $\mathbf{z}\in\mathbb{R}^2$.

2. Let $\mathbb{F}$ and V be as above, and $\Phi(\mathbf{x},\mathbf{y}) = x_1y_1 - x_2y_2$. In this case we have the commutativity identity $\Phi(\mathbf{y},\mathbf{x}) = \Phi(\mathbf{x},\mathbf{y})$ for all $\mathbf{x}$ and $\mathbf{y}$, but if $\mathbf{z}=(1,1)$, or any multiple of the latter, then $\Phi(\mathbf{z},\mathbf{z}) = 0$.

3. Let A be an $n\times n$ matrix over $\mathbb{F}$, and let V be the vector space of all $n\times 1$ column matrices. Define a bilinear form Φ_A on V by the formula

$$\Phi_A(\mathbf{x},\mathbf{y}) = {}^{\mathbf{T}}\mathbf{x}A\mathbf{y} .$$

Examples of this sort appeared frequently in the preceding section (see also Appendix E). Actually, the first two examples are special cases of this construction in which A is given as follows:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad\qquad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

In fact, the following theorem shows that, in principle, the preceding construction gives **all** possible bilinear forms on finite-dimensional vector spaces.

THEOREM VII.7. *Let v be ann-dimensional vector space over $\mathbb{F}$, and let $\mathcal{A} = \{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$ be an ordered basis for V. If Φ is a bilinear form over $\mathbb{F}$, let $[\Phi]_{\mathcal{A}}$ be the $n\times n$ matrix whose (i,j) entry is equal to $\Phi(\mathbf{a}_i,\mathbf{a}_j)$. Then the map sending Φ to $[\Phi]_{\mathcal{A}}$ defines a $1-1$ correspondence between bilinear forms over V and $n\times n$ matrices over $\mathbb{F}$.*

The matrix $[\Phi]_{\mathcal{A}}$ is called *the matrix of Φ with respect to the ordered basis $\mathcal{A}$.*

Proof. *The mapping is* $1-1$. Suppose that we are given two bilinear forms Φ and Ψ such that $\Phi(\mathbf{a}_i, \mathbf{a}_j) = \Psi(\mathbf{a}_i, \mathbf{a}_j)$ for all i and j (this is the condition for $[\Phi]_{\mathcal{A}}$ and $[\Psi]_{\mathcal{A}}$ to be equal). If $\mathbf{v},\ \mathbf{w} \in V$, express these vectors as linear combinations of the basis vectors as follows:

$$\mathbf{v} = \sum_i x_i \mathbf{a}_i \qquad \mathbf{w} = \sum_j y_j \mathbf{b}_j$$

Then by **(Bi–1)** — **(Bi–3)** we have

$$\Phi(\mathbf{v}, \mathbf{w}) = \sum_{i,j} x_i y_j \Phi(\mathbf{a}_i, \mathbf{a}_j) = \sum_{i,j} x_i y_j \Psi(\mathbf{a}_i, \mathbf{a}_j) = \Psi(\mathbf{v}, \mathbf{w})$$

and since $\mathbf{v}$ and $\mathbf{w}$ are arbitrary we have $\Phi = \Psi.\Box$

The mapping is onto. If B is an $n \times n$ matrix and $\mathbf{v},\ \mathbf{w} \in V$ are as in the preceding paragraph, define

$$\mathbf{f}_{B,\mathcal{A}} = \sum_{i,j} x_i y_j b_{i,j} \ .$$

This is well-defined because the coefficients of $\mathbf{v}$ and $\mathbf{w}$ with respect to $\mathcal{A}$ are uniquely determined. The proof that $\mathbf{f}_{B,\mathcal{A}}$ satisfies **(Bi–1)** — **(Bi–3)** is a sequence of routine but slightly messy calculations, and it is left as an exercise. Given this, it follows immediately that B is equal to $[\mathbf{f}_{B\mathcal{A}}]_{\mathcal{A}}.\blacksquare$

CHANGE OF BASIS FORMULA. Suppose we are given a bilinear form Φ on an n-dimensional vector space V over $\mathbb{F}$, and let $\mathcal{A}$ and $\mathcal{B}$ be ordered basis for V. In several contexts it is useful to understand the relationship between the matrices $[\Phi]_{\mathcal{A}}$ and $[\Phi]_{\mathcal{B}}$. The equation relating these matrices are given by the following result:

THEOREM VII.8. *Given two ordered bases $\mathcal{A}$ and $\mathcal{B}$, define a transition matrix by the form*

$$\mathbf{b}_j = \sum_i p_{i,j} \mathbf{a}_i \ .$$

If Φ is a bilinear form on V as above, then we have

$$[\Phi]_{\mathcal{B}} = {}^{\mathbf{T}}P [\Phi]_{\mathcal{A}} \, P \ .$$

Proof. We only need to calculate $\Phi(\mathbf{b}_i, \mathbf{b}_j)$; by the equations above, we have

$$\Phi(\mathbf{b}_i, \mathbf{b}_j) = \Phi\left(\sum_k p_{k,i} \mathbf{a}_k, \sum_m p_{m,j} \mathbf{a}_m\right) =$$

$$\sum_k \left(p_{k,i} \left(\sum_m p_{m,k} \, \Phi(\mathbf{a}_k, \mathbf{a}_m) \right) \right) \ .$$

However, the coefficient of $p_{k,i}$ is just the (k,j) entry of $[\Phi]_{\mathcal{A}} \, P$, and hence the entire summation is just the (i,j) entry of ${}^{\mathbf{T}}P[\Phi]_{\mathcal{A}} \, P$, as claimed.$\blacksquare$

Definition. A bilinear form Φ is *symmetric* if $\Phi(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}$ and $\mathbf{y}$.

THEOREM VII.9. *Let Φ and $\mathcal{A}$ be as in Theorem 7. Then Φ is symmetric if and only if $[\Phi]_{\mathcal{A}}$ is a symmetric matrix.*

Proof. Suppose that Φ is symmetric. Then $\Phi(\mathbf{a}_i, \mathbf{a}_j) = \Phi(\mathbf{a}_j, \mathbf{a}_i)$ for all i and j, and this implies that $[\Phi]_{\mathcal{A}}$ is a symmetric matrix.

Conversely, if $[\Phi]_{\mathcal{A}}$ is symmetric and $\mathbf{v},\ \mathbf{w} \in V$ (the same notation as in Theorem 7), then by Theorem 7 we have

$$\Phi(\mathbf{v}, \mathbf{w}) \ = \ \sum_{i,j} \ ([\Phi]_{\mathcal{A}})_{i,j}\, x_i y_j \qquad\qquad \Phi(\mathbf{w}, \mathbf{v}) \ = \ \sum_{i,j} \ ([\Phi]_{\mathcal{A}})_{j,i}\, x_i y_j \ .$$

Since $[\Phi]_{\mathcal{A}}$ is symmetric, the two summations are equal, and therefore we must have

$$\Phi(\mathbf{y}, \mathbf{x}) \ = \ \Phi(\mathbf{x}, \mathbf{y})$$

for all $\mathbf{x}$ and $\mathbf{y}$.■

We have introduced all of the preceding algebraic machinery in order to prove the following result:

THEOREM VII.10. *Let $\mathbb{F}$ be a field in which $1 + 1 \neq 0$, and let A be a symmetric $n \times n$ matrix over $\mathbb{F}$. Then there is an invertible matrix P such that ${}^{\mathbf{T}}PAP$ is a diagonal matrix.*

This will be a consequence of the next result:

THEOREM VII.11. *Let Φ be a symmetric bilinear form on an n-dimensional vector space V over a field $\mathbb{F}$ for which $1 + 1 \neq 0$. Then there is an ordered basis $\mathbf{v}_1, \ \cdot\ , \mathbf{v}_n$ of V such that $\Phi(\mathbf{v}_i, \mathbf{v}_j) \ = \ 0$ if $i \neq j$ and $\Phi(\mathbf{v}_i, \mathbf{v}_i) \ = \ d_i$ for suitable scalars $d_i \in \mathbb{F}$.*

Proof that Theorem 11 implies Theorem 10. Define a bilinear form Φ_A as in Example 1 above. By construction $[\Phi_A]_{\mathcal{U}} \ = \ A$, where $\mathcal{U}$ is the ordered basis obtained of standard unit vectors. On the other hand, if $\mathcal{V}$ is the ordered basis obtained from Theorem 11, then $[\Phi_A]_{\mathcal{V}}$ is a diagonal matrix. Apply Theorem 8 with $\Phi \ = \ \Phi_A$, $\mathcal{A} \ = \ \mathcal{U}$, and $\mathcal{B} \ = \ \mathcal{V}$.■

Proof of Theorem 11. If $\dim V \ = 1$, the result is trivial. Assume by induction that the result holds for vector spaces of dimension $n - 1$.

CASE 1. Suppose that $\Phi(\mathbf{x}, \mathbf{x}) \ = \ 0$ for all $\mathbf{x}$. Then $\Phi(\mathbf{x}, \mathbf{y}) \ = \ 0$ for all $\mathbf{x}$ and $\mathbf{y}$ because we have

$$\Phi(\mathbf{x}, \mathbf{y}) \ = \ \tfrac{1}{2}\Phi(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) \ - \ \Phi(\mathbf{x}, \mathbf{x}) \ - \ \Phi(\mathbf{y}, \mathbf{y})$$

and consequently $[\Phi]_{\mathcal{A}} \ = \ \mathbf{0}$ for *every* ordered basis $\mathcal{A}$.□

CASE 2. Suppose that $\Phi(\mathbf{v}, \mathbf{v}) \ \neq \ 0$ for some $\mathbf{v}$. Let W be the set of all $\mathbf{x} \in V$ such that $\Phi(\mathbf{x}, \mathbf{v}) \ = \ 0$.[4] We claim that $W \ + \ \mathbb{F} \cdot \mathbf{v} \ = \ V$ and $W \cap \mathbb{F} \cdot \mathbf{v} \ = \ \{\mathbf{0}\}$. — The second assertion is trivial because $\Phi(\mathbf{v}, c \cdot \mathbf{v}) \ = \ 0$ implies that $c \cdot \Phi(\mathbf{v}, \mathbf{v}) \ = \ 0$. Since $\Phi(\mathbf{v}, \mathbf{v}) \ \neq \ 0$, this can only happen if $c \ = \ 0$, so that $c \cdot \mathbf{v} \ = \ \mathbf{0}$. To prove the first assertion, we must observe that for arbitrary $\mathbf{v} \in V$ the vector

$$\Pi(\mathbf{x}) \ = \ \mathbf{x} \ - \ \frac{\Phi(\mathbf{x}, \mathbf{v})}{\Phi(\mathbf{v}, \mathbf{v})}\, \mathbf{v}$$

[4] If Φ is the usual dot product on $\mathbb{R}^n$, then this is the hyperplane through $\mathbf{0}$ that is perpendicular to the line $\mathbf{0v}$.

lies in W (to verify this, compute $\Phi\big(\Pi(\mathbf{x}),\,\mathbf{v}\big)$ explicitly).[5] The conditions on W and $\mathbb{F}\cdot\mathbf{v}$ together with the dimension formulas imply that $\dim W \;=\; n-1$.

Consider the form Ψ obtained by restricting Φ to W; it follows immediately that Ψ is also symmetric. By the induction hypothesis there is a basis $\mathbf{w}_1,\ \cdots\ ,\mathbf{w}_{n-1}$ for W such that $\Phi(\mathbf{w}_i,\mathbf{w}_j) \;=\; 0$ if $i \neq j$. If we adjoin $\mathbf{v}$ to this set, then by the conditions on W and $\mathbb{F}\cdot\mathbf{v}$ we obtain a basis for V. Since $\Phi(\mathbf{v},\mathbf{w}_j)$ is zero for all j by the definition of W, it follows that the basis for V given by $\mathbf{v}$ together with $\mathbf{w}_1,\ \cdots\ ,\mathbf{w}_{n-1}$ will have the desired properties.■

The proof above actually gives and explicit method for finding a basis with the required properties: Specifically, start with a basis $\mathbf{v}_1,\ \cdots\ ,\mathbf{v}_n$ for V. If some $\mathbf{v}_i$ has the property $\Phi(\mathbf{v}_i\mathbf{v}_i) \neq 0$, rearrange the vectors so that the first one has this property. If $\Phi(\mathbf{v}_i,\mathbf{v}_i) \;=\; 0$ for all i, then either $\Phi \;=\; \mathbf{0}$ or else some value $\Phi(\mathbf{v}_i,\mathbf{v}_j)$ is nonzero (otherwise $\Phi = \mathbf{0}$ by Theorem 10). Rearrange the basis so that $\Phi(\mathbf{v}_1,\mathbf{v}_2) \neq 0$, and take a new basis $\{\ \mathbf{v}_i\ \}$ with $\mathbf{v}'_1 \;=\; \mathbf{v}_1+\mathbf{v}_2$ and $\mathbf{v}'_i \;=\; \mathbf{v}_i$ otherwise. Then $\Phi(\mathbf{v}'_1,\mathbf{v}'_1) \neq 0$, and thus in all cases we have modified the original basis to one having this property.

Now we modify $\mathbf{v}'_i$ such that $\mathbf{v}''_1 \;=\; \mathbf{v}'_1$ and $\Phi(\mathbf{v}''_i,\mathbf{v}''_1) = 0$ if $i > 1$. Specifically, if $i \geq 2$ let

$$\mathbf{v}''_i \;=\; \mathbf{v}'_i \;-\; \frac{\Phi(\mathbf{v}'_i,\mathbf{v}'_1)}{\Phi(\mathbf{v}'_1,\mathbf{v}'_1)}\,\mathbf{v}'_1\ .$$

Having done this, we repeat the construction for $\mathbf{w}_1,\ \cdots\ ,\mathbf{w}_{n-1}$ for W with $\mathbf{w}_i = \mathbf{v}''_{i+1}$. When computing explicit numerical examples, it is often convenient to "clear the denominator of fractions" and multiply $\mathbf{v}''_i$ by $\Phi(\mathbf{v}'_1,\mathbf{v}'_1)$. This is particularly true when the matrix $\Phi(\mathbf{v}_i,\mathbf{v}_j)$ are integers (as in Exercise 2 below).

EXERCISES

1. Prove that the map sending bilinear forms to matrices in Theorem 7 is surjective.

2. Find an invertible matrix P such that ${}^{\mathbf{T}}PAP$ is diagonal, where A is the each of the following matrices with real entries:

$$\begin{pmatrix}1&0&1\\0&0&1\\1&1&1\end{pmatrix}\qquad\begin{pmatrix}1&0&1\\0&1&1\\1&1&2\end{pmatrix}\qquad\begin{pmatrix}2&1&3\\1&0&1\\3&1&1\end{pmatrix}\qquad\begin{pmatrix}0&1&0\\0&1&1\\0&1&1\end{pmatrix}\ .$$

3. A symmetric bilinear form Φ on an n-dimensional vector space V over a field $\mathbb{F}$ is said to be *nondegenerate* if for each nonzero $\mathbf{x} \in V$ there is some $\mathbf{y} \in V$ such that $\Phi(\mathbf{x},\mathbf{y}) \neq 0$. Given an ordered basis $\mathcal{A}$ for V, show that Φ is nondegenerate if and only if the matrix $[\Phi]_{\mathcal{A}}$ is invertible. [*Hint:* Suppose that $\mathbf{x}$ satisfies $B\mathbf{x} \;=\; \mathbf{0}$, where B is the matrix in the previous sentence, and let $\mathbf{v} \;=\; \sum_i\ x_i\mathbf{a}_i$. If $\mathbf{w} \;=\; \sum_j\ y_j\mathbf{z}_j$, explain why ${}^{\mathbf{T}}\mathbf{y}B\mathbf{x} \;=\; \Phi(\mathbf{x},\mathbf{y})$ and how this is relevant.]

[5] If Φ is the ordinary dot product, then $\Pi(\mathbf{x})$ is the foot of the perpendicular dropped from $\mathbf{x}$ to the plane determined by W, and hence $\mathbf{0}\Pi(\mathbf{x})$ is perpendicular to W.

4. Projective classification of hyperquadrics

A standard exercise in plane and solid analytic geometry is the classification of conics and quadrics up to changes of coordinates given by rotations, reflections and translations. Stated differently, the preceding is the classification up to finding a rigid motion sending on to the other. An account of the classification for arbitrary dimensions appears on pages 257–262 of Birkhoff and MacLane, *Survey of Modern Algebra* ($3^{\rm rd}$ Ed.). A related classification (up to finding an affine transformation instead of merely a rigid motion) is discussed in Exercise 4 below. In this section we are interested in the corresponding projective problem involving projective hyperquadrics and (projective) collineations.

Throughout this section we assume that $\mathbb{F}$ is a field in which $1+1 \neq 0$. Furthermore, if $\Sigma \subset \mathbb{FP}^n$ is a hyperquadric, then we shall use $\mathsf{SingSet}(\Sigma)$ to denote its subset of singular points.

We shall begin with an important observation.

THEOREM VII.12. *Let g be a projective collineation of $\mathbb{FP}^n$. Then a subset $\Sigma \subset \mathbb{FP}^n$ is a hyperquadric if and only if $g[\Sigma]$ is. Furthermore, the singular sets of these hyperquadrics satisfy*

$$g\,[\,\mathsf{SingSet}(\Sigma)\,] \;=\; \mathsf{SingSet}\,(\,T[\Sigma]\,)$$

and if $\mathrm{Tang}_X(\Sigma)$ *denotes the tangent hyperplane to Σ at a nonsingular point X, then*

$$g\,[\,\mathrm{Tang}_X(\Sigma)\,] \;=\; \mathrm{Tang}_{g(X)}\,(\,T[\Sigma]\,)\ .$$

Proof. Let A be a symmetric $(n+1)\times(n+1)$ matrix which defines the hyperquadric Σ.

According to Theorem VI.14, there is an invertible linear transformation C of $\mathbb{F}^{n+1,1}$ such that $T\,(\mathbb{F}\cdot\xi) \;=\; \mathbb{F}\cdot C(\xi)$ for all nonzero vectors $\xi \in \mathbb{F}^{n+1,1}$. Let B be the matrix of C in the standard basis. Then X lies in $T[\Sigma]$ if and only if $T^{-1}(X)$ lies in Σ. If ξ is a set of homogeneous coordinates for X, then the conditions in the preceding sentence are equivalent to

$$^{\mathbf{T}}\xi\,^{\mathbf{T}}B^{-1}AB^{-1}\xi \;-\; 0$$

and the displayed equation is equivalent to saying that X lies on the hyperquadric associated to the (symmetric) matrix $^{\mathbf{T}}B^{-1}AB^{-1}$.

To check the statement about singular points, note that a point X lies on $\mathsf{SingSet}(\Sigma)$ if and only if X has homogeneous coordinates ξ such that $^{\mathbf{T}}\xi A \;=\; \mathbf{0}$, and the latter is equivalent to

$$^{\mathbf{T}}\xi\,^{\mathbf{T}}B\,^{\mathbf{T}}B^{-1}AB^{-1} \;=\; \mathbf{0}$$

which in turn is equivalent to

$$^{\mathbf{T}}(B\xi)\cdot\left(^{\mathbf{T}}B^{-1}AB^{-1}\right) \;=\; \mathbf{0}\ .$$

To check the statement on tangent hyperplanes, note that Y lies on the tangent hyperplane to Σ at X if and only if there are homogeneous coordinates ξ for X and η for Y such that $^{\mathbf{T}}\xi A\eta \;=\; \mathbf{0}$, and the latter is equivalent to

$$^{\mathbf{T}}\xi\,^{\mathbf{T}}B\,^{\mathbf{T}}B^{-1}AB^{-1}B\eta \;=\; \mathbf{0}$$

which in turn is equivalent to

$$^{\mathbf{T}}(B\xi)\cdot\left(tp\,B^{-1}AB^{-1}\right)\eta) \;=\; \mathbf{0}\ .$$

The latter is equivalent to saying that $T(Y)$ is in the tangent hyperplane to $T[\Sigma]$ at $T(X)$.■

Definition. Two hypequadrics Σ andΣ' are *projectively equivalent* if there is a projective collineation T such that $T[\Sigma] = \Sigma'$. We sometimes write this relation as $\Sigma \sim \Sigma'$. It is clearly an equivalence relation, and the main goal of this section is to understand this relation when $\mathbb{F}$ is the real or complex numbers.

We shall first describe some necessary and sufficient conditions for the projective equivalence of hyperquadrics.

THEOREM VII.13. *Let Σ be a hyperquadric in $\mathbb{FP}^n$, and let T be a projective collineation of $\mathbb{FP}^n$. Then the following hold:*

(i) The dimensions of the geometrical subspaces of singular points of Σ and $T[\Sigma]$ must be equal.

(ii) If Σ contains no geometrical subspace of dimension r, then neither does $T[\Sigma]$.

Proof. (i) By definition, $\mathsf{SingSet}(\Sigma)$ is the set of all X whose homogeneous coordinates ξ satisfy ${}^{\mathbf{T}}\xi A = \mathbf{0}$, and hence $\mathsf{SingSet}(\Sigma)$ is a geometrical subspace. Now Theorem 12 implies that $T[\mathsf{SingSet}(\Sigma)] = \mathsf{SingSet}T[\Sigma]$, and hence

$$\dim \mathsf{SingSet}(\Sigma) = \dim T[\mathsf{SingSet}(\Sigma)] = \dim \mathsf{SingSet}T[\Sigma] \;.\square$$

(ii) Suppose $Q \subset T[\Sigma]$ is an r-dimensional geometrical subspace. Since T^{-1} is also a projective collineation, the set

$$T^{-1}[Q] \subset T^{-1}\left[T[\Sigma]\right] = \Sigma$$

is also an r-plane.■

THEOREM VII.14. *Suppose that Σ and Σ' are hyperquadrics which are defined by the symmetric matrices A and B respectively. Assume that there is an invertible matrix C and a nonzero constant k such that $B = {}^{\mathbf{T}}CAC$. Then Σ and Σ' are projectively equivalent.*

Proof. Let T be the projective collineation defined by C^{-1}, and if $X \in \mathbb{FP}^n$ let ξ be a set of homogeneous coordinates for X. Then by Theorem 12 we have

$$T[\Sigma] = \{\, \mathbf{x} \mid {}^{\mathbf{T}}\xi{}^{\mathbf{T}}CAC\xi = 0 \,\} = \{\, \mathbf{x} \mid {}^{\mathbf{T}}\xi(k^{-1}B)\xi = 0 \,\} = \{\, \mathbf{x} \mid k^{-1}\left({}^{\mathbf{T}}\xi B\xi\right) = 0 \,\} = \Sigma' \;.\blacksquare$$

NOTATION. Let D_r be the $n \times n$ diagonal matrix $(n \geq r)$ with ones in the first r entries and zeros elsewhere, and let $D_{p,q}$ denote the $n \times n$ diagonal matrix $(n \geq p+q)$ with ones in the first entries, (-1)'s in the next q entries, and zeros elsewhere.

REMARKS. 1. If A is a symmetric matrix over the complex numbers, then for some invertible matrix P the product ${}^{\mathbf{T}}PAP$ is D_r for some r. For Theorem 10 guarantees the existence of an invertible matrix P_0 such that $A_1 = {}^{\mathbf{T}}P_0AP_0$ is diagonal. Let P_1 be the diagonal matrix whose entries are square roots of the corresponding nonzero diagonal entries of A_1, and ones in the places where A_1 has zero diagonal entries. Then the product $P = P_0P_1^{-1}$ has the desired properties. This uses the fact that every element of the complex numbers $\mathbb{C}$ has a square root in $\mathbb{C}$, and in fact the same argument works in every field $\mathbb{F}$ which is closed under taking square roots.

2. If A is a symmetric matrix over the complex numbers, then for some invertible matrix P the product ${}^{\mathbf{T}}PAP$ is $D_{p,q}$ for some p and q. As in the preceding example, choose an invertible matrix P_0 such that $A_1 = {}^{\mathbf{T}}P_0AP_0$ is diagonal. Let P_1 be the diagonal matrix whose entries are square roots of the **absolute values** of the corresponding nonzero diagonal entries of A_1, and ones in the places where A_1 has zero diagonal entries. Then the product $P = P_0P_1^{-1}$ has the desired properties. The need for more complicated matrices arises because over $\mathbb{R}$ one only has square roots of nonnegative numbers, and if $x \in \mathbb{R}$ then either x or $-x$ is nonnegative.

The preceding remarks and Theorems 12–14 yield a complete classification of hyperquadrics in $\mathbb{FP}^n$ up to projective equivalence if $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. We shall start with the complex case, which is easier.

THEOREM VII.15. *Let $\Gamma_r \subset \mathbb{CP}^n$ defined by the matrix D_r described above. Then every nonempty hyperquadric in $\mathbb{CP}^n$ is projectively equivalent to Γ_r for some uniquely determined value of r.*

Proof. By Remark 1 above and Theorem 14, we know that Σ is projectively equivalent to Γ_r for some r. It suffices to show that if Γ_r and Γ_s are projectively equivalent then $r = s$.

By the preceding results we know that $\dim \mathsf{SingSet}(\Gamma_r)$ is the dimension of the subspace of all X whose homogeneous coordinates ξ satisfy ${}^{\mathbf{T}}\xi D_r = \mathbf{0}$, and the dimension of that subspace is equal to $n - r + 1$. Therefore, if Γ_r and Γ_s are projectively equivalent then we must have $n - r + 1 = n - s + 1$, which implies that $r = s$, so there is only one such hyperquadric that can be equivalent to Σ and thus uniqueness follows.■

The preceding argument goes through if $\mathbb{C}$ is replaced by an arbitrary field $\mathbb{F}$ which is closed under taking square roots.

Over the real numbers, the classification is somewhat more complicated but still relatively simple.

THEOREM VII.16. *Let $\Gamma_{p,q} \subset \mathbb{RP}^n$ defined by the diagonal matrix $D_{p,q}$ described above. Then every nonempty hyperquadric in $\mathbb{RP}^n$ is projectively equivalent to $\Gamma_{p,q}$ for some uniquely determined values of p and q such that $p \geq q$.*

Proof. As in the proof of the preceding theorem, by Theorem 14 and Remark 1 we know that an arbitrary projective hyperquadric is projectively equivalent to $\Gamma_{p,q}$ for some p and q. This hyperquadric is represented by $D_{p,q}$; if we permute the homogeneous coordinates, we see that $\Gamma_{p,q}$ is projectively equivalent to the hyperquadric defined by the matrix $-D_{q,p}$, and since the negative of this matrix defines the same hyperquadric it follows that $\Gamma_{p,q}$ is projectively equivalent to $\Gamma_{q,p}$. Since either $p \geq q$ or $q \geq p$, it follows that every hyperquadric is projectively equivalent to $\Gamma_{u,v}$ for some $u \geq v$.

To complete the proof, it will suffice to show that if $\Gamma_{p,q}$ is projectively equivalent to $\Gamma_{u,v}$ where $p \geq q$ and $u \geq v$, then $p+q = u+v$ and $p = u$. To see the first equality, note that the dimension of $\mathsf{SingSet}(\Gamma_{a,b})$ is equal to $n - (a+b) + 1$ by the argument in the preceding theorem, and as in that proof we conclude that $p + q = u + v$.

To see the second equality, we shall characterize the integer p in $\Gamma_{p,q}$ as follows.

(‡) *The hyperquadric $\Gamma_{p,q}$ contains a geometric subspace of dimension $n-p$ but no such subspace of higher dimension.*

This and the second part of Theorem 13 will combine to prove that if $\Gamma_{p,q}$ is projectively equivalent to $\Gamma_{u,v}$ where $p \geq q$ and $u \geq v$, then we also have $p = u$.

An explicit geometrical subspace S of dimension $N-p$ is given by the equations

$$\begin{array}{cc} x_i - x_{p+i} = 0 & 1 \leq i \leq q, \\ x_i = 0 & q \leq i \leq p\ . \end{array}$$

Consider the geometrical subspace T defined by

$$x_{p+1} = x_{p+2} = \cdots = x_{n+1} = 0\ .$$

This geometrical subspace is $(p-1)$-dimensional. Furthermore, if $X \in T \cap \Sigma$ has homogeneous coordinates $(x_1, \cdots, x_{n+1})$ we have $x_i = 0$ for $i > p$, so that

$$\sum_{i \leq p} x_i^2 = 0\ .$$

The latter implies that $x_i = 0$ for $i \leq p$, and hence it follows that $x_i = 0$ for **all** i; this means that *the intersection $T \cap \Sigma$ is the empty set.*

Suppose now that $S' \subset \Sigma$ is a geometrical subspace of dimension $\geq n-p+1$. Then the addition law for dimensions combined with $\dim(S' \star T) \leq n$ shows that $S' \cap T \neq \varnothing$, and since $S' \subset \Sigma$ we would also have $\Sigma \cap T \neq \varnothing$. But we have shown that the latter intersection is empty, and hence it follows that Σ cannot contain a geometrical subspace of dimension greater than $(n-p)$, which is what we needed to show in order to complete the proof.■

COMPUTATIONAL TECHNIQUES. Over the real numbers, there is another standard method for finding an equivalent hyperquadric defined by a diagonal matrix. Specifically, one can use the following diagonalization theorem for symmetric matrices to help find a projective collineation which takes a given hyperquadric to one of the given type:

Let A be a symmetric matrix over the real numbers. Then there is an orthogonal matrix P (one for which ${}^{\mathbf{T}}P = P^{-1}$) such that ${}^{\mathbf{T}}PAP$ is a diagonal matrix. Furthermore, if λ_i is the i^{th} entry of the diagonal matrix, then the i^{th} column of P is an eigenvector of A whose associated eigenvalue is equal to λ_i.■

This statement is often called the **Fundamental Theorem on Real Symmetric Matrices**, and further discussion appears on pages 51–52 of the following online notes:

`http://math.ucr.edu/~res/math132/linalgnotes.pdf`

If we combine the Fundamental Theorem on Real Symmetric Matrices with other material from this section, we see that the construction of a projective collineation taking the hyperquadric Σ_A defined by A to a hyperquadric defined by an equation of the form

$$\Sigma_i\ d_i x_i^2 = 0$$

reduces to finding the eigenvalues and eigenvectors of A. This approach is probably the most effective general method for solving problems like those in Exercise 3 below.

SPECIALIZATION TO THE REAL PROJECTIVE PLANE. We shall conclude this section by restating a special case of Theorem 16 that plays a crucial role in Section 6.

THEOREM VII.17. *All nonempty nonsingular conics in* $\mathbb{RP}^2$ *are projectively equivalent. In fact, they are equivalent to the affine unit circle which is defined by the homogeneous coordinate equation* $x_1^2 + x_2^2 - x_3^2 = 0$.

Proof. We must consider all $\Gamma_{p,q}$ with $p \geq q$ and $p + q = 3$ (this is the condition for the singular set to be empty). The only possibilities for (p, q) are $(2, 1)$ and $(3, 0)$. However, $\Gamma_{3,0}$ — the set of points whose homogeneous coordinates satisfy $x_1^2 + x_2^2 + x_3^2 = 0$ — is empty, so there is a unique possibility and it is given by $\Gamma_{2,1}$, which is the affine unit circle.■

EXERCISES

1. For each projective quadric in Exercise VII.1.3, determine the quadric in $\mathbb{RP}^3$ to which it is projectively equivalent.

2. Show that the number of projective equivalence classes of hypequadrics in $\mathbb{RP}^n$ is equal to $\frac{1}{4}(n+2)(n+4)$ if n is even and $\frac{1}{4}(n+3)^2$ if n is odd.

3. For each of the examples below, find a projective collineation of $\mathbb{RP}^2$ that takes the projectivizations of the following affine conics into the unit circle (with affine equation $x^2 + y^2 = 1$).

(i) The hyperbola $xy = 4$.

(ii) The parabola $y = x^2$.

(iii) The ellipse $4x^2 + 9y^2 = 36$.

(iv) The hyperbola $4x^2 - 9y^2 = 36$.

4. (a) What should it mean for two **affine** hyperquadrics in $\mathbb{R}^n$ to be affinely equivalent?

(b) Prove that every affine hyperquadric in $\mathbb{R}^n$ is equivalent to one defined by an equation from the following list:

$$x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 = 0 \qquad (r \leq n)$$

$$x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 + 1 = 0 \qquad (r \leq n)$$

$$x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 + x_{r+1} = 0 \qquad (r < n)$$

See Birkhoff and MacLane, pp. 261–264, or Section V.2 of the online notes

`http://math.ucr.edu/~res/math132/linalgnotes.pdf`

for further information on this topic.

5. Duality and projective hyperquadrics

In this section we shall show that the duality properties for geometrical subspaces of coordinate projective spaces are part of a larger pattern of duality which includes hyperquadrics. As in most other sections of these notes, $\mathbb{F}$ will denote a (commutative) field in which $1 + 1 \neq 0$.

Definition. A *hypersurface of the second class* in $(\mathbb{FP}^n)^*$ is the set $\mathfrak{F}$ of all hyperplanes H whose homogeneous coordinates θ satisfy a homogeneous quadratic equation

$$\theta B^{\mathbf{T}}\theta \;=\; 0.$$

If we write $\theta \;=\; (u_1, \;\cdots\;, u_{n+1})$ and B has entries $b_{i,j}$, this is equivalent to the scalar equation $\sum_{i,j}\; b_{i,j}u_iu_j \;=\; 0$.

The dual of a tangent line is a *cotangent* $(n-2)$-*subspace* V contained in a hyperplane H which belongs to the hypersurface $\mathfrak{F}$. The cotangent condition becomes an assertion that *exactly one of the following two statements is valid*:

- (i) H is the only hyperplane in $\mathfrak{F}$ containing V.
- (ii) Every hyperplane in $\mathfrak{F}$ contains V.

In the first case we say that H is a *nonsingular* hyperplane in $\mathfrak{F}$, and in the second we say that H is a *singular* hyperplane in $\mathfrak{F}$.

By duality the set of all cotangent $(n-2)$-spaces at a nonsingular hyperplane is the set of all $(n-2)$-spaces through a point called the *point of contact* of $\mathfrak{F}$ at H. This point has homogeneous coordinates $B^{\mathbf{T}}\theta$.

Nonsingular hyperquadrics and nonsingular hypersurfaces of the second class satisfy the following useful interrelationship:

THEOREM VII.18. (i) *The set of all hyperplanes* $\mathbf{T}\Sigma$ *tangent to a nonsingular hyperquadric* Σ *is a hypersurface of the second class.*

(ii) *The set of all points of contact* $\mathbf{K}\mathfrak{F}$ *to a nonsingular hypersurface* $\mathfrak{F}$ *of the second class is a nonsingular hyperquadric.*

(iii) *In the setting of the preceding two statements, we have* $\mathbf{KT}\Sigma \;=\; \Sigma$ *and* $\mathbf{TK}\mathfrak{F} \;=\; \mathfrak{F}$.

Proof. (i) Suppose Σ is defined as the set of all X whose homogeneous coordinates satisfy ${}^{\mathbf{T}}\xi A\xi \;=\; 0$. We claim that H lies in $\mathbf{T}\Sigma$ if and only if its homogeneous coordinates θ satisfy $\theta A^{-1}\,{}^{\mathbf{T}}\theta \;=\; 0$.

Suppose that $H \in \mathbf{T}\Sigma$. Let X be a point such that H is the tangent hyperplane to X, and let θ and ξ be homogeneous coordinates for H and X respectively. Then we have $\theta \;=\; {}^{\mathbf{T}}\xi A$, and hence

$$\theta A^{-1}\,{}^{\mathbf{T}}\theta \;=\; \left({}^{\mathbf{T}}\xi A\right)\theta A^{-1}(A\xi) \;=\; {}^{\mathbf{T}}\xi A\xi \;=\; 0$$

which is what we wanted to prove.

Conversely, suppose that homogeneous coordinates θ for H satisfy the equation

$$\theta A^{-1}\,{}^{\mathbf{T}}\theta \;=\; 0\;.$$

Let $\xi = A^{-1\,\mathbf{T}}\theta$, and let X have homogeneous coordinates ξ. Then $\theta = {}^{\mathbf{T}}\xi A$ and

$$ {}^{\mathbf{T}}\xi A\xi \;=\; {}^{\mathbf{T}}\xi A\,A^{-1}\,A\xi \;=\; \theta A^{-1}\,{}^{\mathbf{T}}\theta \;=\; 0 $$

so that H lies in $\mathbf{T}\Sigma$.

Finally, note that $\mathbf{T}\Sigma$ is nonsingular because it is defined by the invertible matrix A^{-1}.$\square$

(ii) The statement about $\mathbf{K}\mathfrak{F}$ follows by duality. It remains to show that $\mathbf{KT}\Sigma = \Sigma$. Howver, if Σ is defined by the invertible matrix A, then $\mathbf{T}\Sigma$ is defined by the inverse matrix A^{-1}, and therefore by duality it follows that $\mathbf{KT}\Sigma$ is defined by the matrix

$$ \left(A^{-1}\right)^{-1} \;=\; A $$

and hence it must be equal to Σ.

Finally, the assertion $\mathbf{TK}\mathfrak{F} = \mathfrak{F}$ follows by duality.$\blacksquare$

Extending duality to nonsingular hyperquadrics. The preceding theorem implies the following general principle:

AUGMENTED PRINCIPLE OF DUALITY. *A statement about coordinate projective n-spaces over fields remains true if — in addition to the previously specified interchanges involving geometrical subspaces — one interchanges the phrases* point on a nonsingular hyperquadric *and* tangent hyperplane to a nonsingular hyperquadric.$\blacksquare$

Important examples of this extended dualization will be given in the next (and final) section of these notes.

EXERCISES

1. Find the equations defining the tangent lines to the projectivizations of the following affine conics:

(i) The parabola $y^2 = 4ax$.

(ii) The ellipse $a^2x^2 + b^2y^2 = a^2b^2$.

(iii) The hyperbola $a^2x^2 - b^2y^2 = a^2b^2$.

(iv) The hyperbola $xy = a$.

2. Find the equation defining the conic in $\mathbb{RP}^2$ whose tangent lines satisfy the equation

$$ u_1^2 \;-\; 2u_1u_2 \;+\; u_2^2\,2u_2u_3 \;+\; 2u_1u_3 \;+\; u_3^2 \;=\; 0\;. $$

[*Hint:* Look at the proof of Theorem 18.]

3. Write out the plane dual to the following statements about conics in the projective plane $\mathbf{P}$:

(i) At the points X and Y on the nonsingular conic Γ, the respective tangent lines L and M meet at a point Z.

(ii) No three points of the nonsingular conic Γ are collinear.

(iii) There are two lines in the (projective) plane $\mathbf{P}$ that are tangent to both of the nonsingular conics Γ_1 and Γ_2.

6. Conics in the real projective plane

Projective conics have a great many interesting properties, the most famous of which is Pascal's Theorem (see Theorem 24 below). A thorough discussion of projective conics appears in Coolidge, *A History of the Conic Sections and Quadric Surfaces*. In this section we shall limit ourselves to proving a few of the more important and representative theorems in the subject.

Throughout this section we shall be considering coordinate projective planes over a fixed field $\mathbb{F}$ in which $1 + 1 \neq 0$. We shall also assume that $\mathbb{F}$ is not isomorphic to $\mathbb{Z}_3$ after Theorem 22. Of course, this means that all the results in this section are valid in the real and complex projective planes.

THEOREM VII.19. *Given any five points in $\mathbb{FP}^2$, no three of which are collinear, there is a unique conic containing them. Furthermore, this conic is nonsingular.*

Proof. Let $A,\ B,\ C,\ D,\ E,\ V$ be five points, no three of which are collinear. We shall first prove the result in a special case and then prove that it holds more generally.

CASE 1. Suppose that homogeneous coordinates $\alpha,\ \beta,\ \gamma,\ \delta$ for $A,\ B,\ C,\ D$ are given by standard values:

$$\alpha = \begin{pmatrix}1\\0\\0\end{pmatrix} \qquad \beta = \begin{pmatrix}0\\1\\0\end{pmatrix} \qquad \gamma = \begin{pmatrix}0\\0\\1\end{pmatrix} \qquad \delta = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

Since no three of the five points are collinear, homogeneous coordinates for V are given by $a_0\alpha+b_0\beta+c_0\gamma$, where none of $a_0,\ b_0,\ c_0$ is equal to zero. Dividing by a_0, we can find homogeneous coordinates ψ for V such that $\psi = \alpha + b\beta + c\gamma$. Since $V \notin AD$, the scalars b and c must be distinct.

Suppose there is a conic Γ containing $A,\ B,\ C,\ D,\ E,\ V$ and that it is defined by the symmetric 3×3 matrix P. We need to determine the entries $p_{i,j}$ from the equations ${}^{\mathbf{T}}\xi P \xi = 0$, which should hold for $\xi - \alpha,\ \beta,\ \gamma,\ \delta,\ \psi$. If $\xi = \alpha$, then direct substitutions implies that $p_{1,1} = 0$. Likewise, if we make the substitutions $\xi = \beta$ and $\xi = \gamma$ we find that $p_{2,2} = p_{3,3} = 0$. Continuing in this manner, if we make the substitution $\xi = \delta$ and use the previously derived values for the diagonal entries, we find that

$$2 \cdot (p_{1,2} + p_{2,3} + p_{1,3}) = 0$$

and if we follow this with the substitution $\xi = \psi$ we also obtain the equation

$$2 \cdot (b\,p_{1,2} + bc\,p_{2,3} + c\,p_{1,3}) = 0 .$$

Thus the entries of the symmetric matrix P satisfy the following conditions:

$$p_{1,1} = p_{2,2} = p_{3,3} = 0$$

$$p_{1,2} = \frac{(1-b)c}{b-c}\,p_{2,3} \qquad\qquad p_{1,3} = \frac{(1-c)b}{c-b}\,p_{2,3}$$

Therefore the coefficients of P are uniquely determined up to a scalar multiple, and it follows that there is at most one conic containing the given five points. On the other hand, if we set

$$P \;=\; \begin{pmatrix} 0 & (1-b)c & b(c-1) \\ (1-b)c & 0 & b-c \\ b(c-1) & b-c & 0 \end{pmatrix}$$

then the preceding calculations show that the given five points lie in the conic defined by P.

To prove that the conic is nonsingular, it suffices to show that the determinant of the matrix P defined above is nonzero. First note that $V \notin CD$ and $V \notin BD$ imply $b \neq 1$ and $c \neq 1$ respectively. Thus the determinant of P, which is

$$2\,(1-b)c(b-c)b(b-1)$$

must be nonzero, proving that the conic is nonsingular.□

Case 2. Suppose that $A,\ B,\ C,\ D,\ E,\ V$ are arbitrary. By the Fundamental Theorem of Projective Geometry, there is a projective collineation Φ such that

$$\Phi(A),\ \Phi(B),\ \Phi(C),\ \Phi(D),\ \Phi(E),\ \Phi(V)$$

satisfy the conditions of Case 1. Let Γ_0 be the unique nonsingular conic given by Case 1; then $\Gamma \;=\; \Phi^{-1}[\Gamma_0]$ is a nonsingular conic containing $A,\ B,\ C,\ D,\ E,\ V$ by Theorem 12, proving existence. To show uniqueness, suppose that Γ' is another conic containing the given five points; then $\Phi[\Gamma']$ is a conic containing $\Phi(A),\ \Phi(B),\ \Phi(C),\ \Phi(D),\ \Phi(E),\ \Phi(V)$ and therefore by Case 1 we have $\Phi[\Gamma'] \;=\; \Gamma_0$. Consequently, we have $\Gamma' \;=\; \Phi^{-1} \circ \Phi[\Gamma'] \;=\; \Phi^{-1}[\Gamma_0] \;=\; \Gamma$.■

If $\mathbb{F}$ is not isomorphic to $\mathbb{Z}_3$ then there is a converse to the preceding theorem; namely, every nonsingular conic contains at least five points (see Theorem 21). In fact, no three of these points can be collinear, for the noncollinearity of three arbitrary points on a nonsingular conic will be a consequence of the next result.

Theorem VII.20. *No three points on a nonsingular conic are collinear.*

Proof. Let $A,\ B,\ C$ be three collinear points. We claim that every conic containing all three of them is singular. By the Fundamental Theorem of Projective Geometry and Theorem 12, it suffices to consider the case in which homogeneous coordinates α and β for A and B are the first two unit vectors in $\mathbb{F}^{3,1}$.

By Theorem V.7, homogeneous coordinates γ for C may be chosen so that $\gamma \;=\; \alpha + c\,\gamma$, where $c \neq 0$. If the conic $\Gamma$is defined by the symmetric 3×3 matrix P, then computations like those of Theorem 19 imply that $p_{1,1} \;=\; p_{2,2} \;=\; 0$ and $2c\,p_{1,2} \;=\; 0$. Thus P has the following form:

$$P \;=\; \begin{pmatrix} 0 & 0 & p_{1,3} \\ 0 & 0 & p_{2,3} \\ p_{1,3} & p_{2,3} & p_{3,3} \end{pmatrix}$$

However, direct computation shows that such a matrix is not invertible, and therefore the conic Γ is singular by Theorem 5.■

Here is the other result we need to establish a converse to Theorem 19:

Theorem VII.21. *Let Γ be a nonempty conic in $\mathbb{FP}^2$ containing at least one nonsingular point, and assume that the field $\mathbb{F}$ contains at least n distinct elements. Then Γ contains at least $(n+1)$ distinct points.*

In particular, if $\mathbb{F}$ is not isomorphic to $\mathbb{Z}_3$, then Γ contains at least five distinct points (note that $\mathbb{F}$ cannot be isomorphic to $\mathbb{Z}_2$ because we are assuming that $1+1 \neq 0$ in $\mathbb{F}$).

Proof. Let $X \in \Gamma$ be a nonsingular point, and let L be the tangent line through X. Then there are at least n other lines through X, say $L_1, \cdots, L_n$. Since each L_i is not a tangent line and $X \in L_i \cap \Gamma$, there must be a second point $X_i \in L_i \cap \Gamma$.

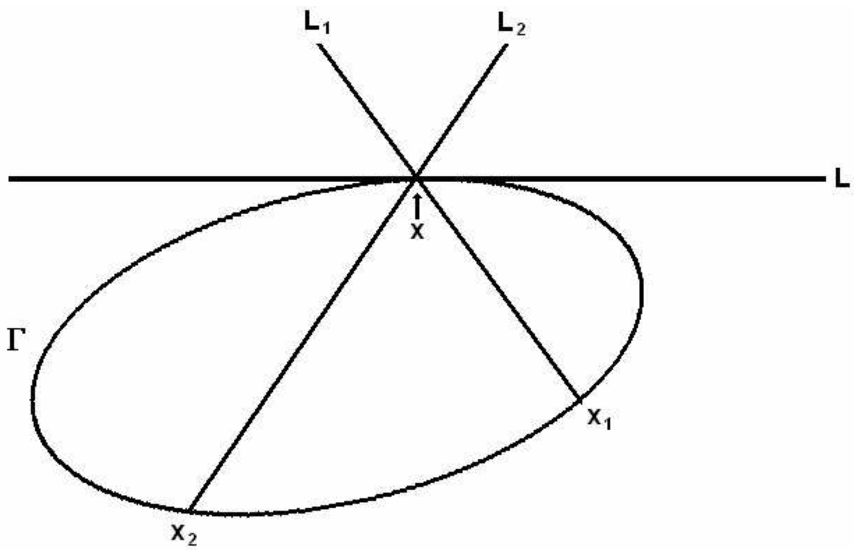


Figure VII.1

If $i \neq j$, then $X_i \neq X_j$ because otherwise L_i and L_j would have two points in common and we know these lines are distinct. Therefore the points $X, X_1, \cdots, X_n$ must be distinct points of Γ.■

A synthetic approach to conics

The theorem above give an incidence-theoretic characterization of nonsingular conics and suggest that synthetic methods might be useful in the study of conics. The next two theorems give a completely synthetic characterization of nonsingular conics due to J. Steiner.[6]

From this point on, unless stated otherwise, we shall assume that the field $\mathbb{F}$ is not isomorphic to $\mathbb{Z}_3$.

Theorem VII.22. *Let A and B be distinct points in $\mathbb{FP}^2$, and let Φ be a projective collineation of $\mathbb{FP}^2$ sending A to B. Then*

$$\mathbf{K} = \{X \in \mathbb{FP}^2 \mid X = A \quad \text{or} \quad X \in \Phi(L) \cap L \text{ for some line } L \text{ through } A\}$$

is a conic. (Notice that $B \in \mathbf{K}$, for we may take $L = AB$ in the definition).

[6]Jakob Steiner (1796–1863) is known for his work on projective geometry from a strongly synthetic viewpoint and for results in other branches of geometry.

Proof. Let P be an invertible 3×3 such that if ξ is a set of homogeneous coordinates for X, then $\Phi(X) = \mathbb{F} \cdot P\xi$. Also, let α be a set of homogeneous coordinates for A, and let β be a set of homogeneous coordinates for B such that $\beta = P \cdot \alpha$.

We need to find a homogeneous quadratic equation which defines $\mathbf{K}$. By Exercise V.1.5, if L is a line in $\mathbb{FP}^2$ and has homogeneous coordinates λ, then the line $\Phi[L]$ has homogeneous coordinates λP^{-1}. Thus $X \in \mathbf{K}$ if and only if $X = A$ or its homogeneous coordinates ξ satisfy

$$\xi = {}^{\mathbf{T}}\big(\lambda P^{-1}\big) \times {}^{\mathbf{T}}\lambda$$

for some line L whose homogeneous coordinates λ satisfy $\lambda \cdot \alpha = 0$. Equivalently, we have $X \in \mathbf{K}$ if and only if $X = A$ or

$$\lambda P^{-1}\xi = \lambda \cdot \xi = \lambda \cdot \alpha = 0 .$$

It follows that $X \in \mathbf{K}$ if and only if α, ξ and $p^{-1}\xi$ are linearly dependent (the case $X \neq A$ is immediate from the preceding three equations, while the case $X = A$ is trivial). Since P defines an invertible linear transformation, the vectors α, ξ and $P^{-1}\xi$ are linearly independent if and only if $\beta = P \cdot \alpha$, $P \cdot \xi$ and $\xi = P P^{-1}\xi$ are linearly independent. The linear dependence of the latter is in turn equivalent to the vanishing of the determinant $[\xi,\ P\xi,\ \beta]$. But the latter expression is a homogeneous quadratic polynomial in the entries of ξ and hence it is the defining equation of a conic.■

Conversely, every nonsingular conic is defined by a projective collineation as in Theorem 22.

THEOREM VII.23. (Steiner) *Let Γ be a nonsingular conic in $\mathbb{FP}^2$ containing at least five distinct points, and let A and B be distinct points of Γ. Then there is a projective collineation Φ of $\mathbb{FP}^2$ sending A to B such that*

$$\Gamma = \{X \in \mathbb{FP}^2 \mid X = A \quad \text{or} \quad X \in \Phi(L) \cap L \text{ for some line } L \text{ through } A\} .$$

Proof. Let X, Y, Z be three points of Γ which are distinct from A and B. By Theorem 20, no three of the points A, B, X, Y, Z are collinear. Thus there is a unique projective collineation Φ sending A to B and X, Y, Z to themselves. By Theorem 22, the points A, B and the collineation Φ determine a conic Γ' defined by the formula above. By construction the three points X, Y, Z lie on Γ', and therefore $\Gamma = \Gamma'$ by Theorem 12.■

NOTATION. If Γ is a conic and $A,\ B \in \Gamma$, then the collineation Φ of Theorem 23 is called a *Steiner collineation* associated to A, B and Γ. We note that this collineation is **not** unique, for different choices of the three points X, Y, Z yield different collineations.

Conics and inscribed polygons

Definition. Let $P_1, \cdots, P_n$ be $n \geq 3$ points in $\mathbb{FP}^2$ such that no three are collinear. The **simple (projective) n-gon** $P_1 \cdots P_n$ is defined to be

$$P_1P_2 \cup \cdots \cup P_{n-1}P_n \cup P_nP_1 .$$

Dually, if $L_1, \cdots, L_n$ is a set of $n \geq 3$ lines such that no three are concurrent, the dual of a simple n-gon is the finite set of points determined by the intersections $L_i \cap L_{i+1}$ and $L_n \cap L_1$ (*i.e.*, a set of n points such that no three are collinear), and the union of the lines is the simple n-gon determined by these n points.

The following result due to B. Pascal[7] is one of the most celebrated theorems in projective geometry:

Theorem VII.24. (Pascal's Theorem) *Suppose that Γ is a nonsingular conic in $\mathbb{FP}^2$ and let the simple hexagon $A_1 \cdots A_6$ be* inscribed *in Γ (in other words, $A_i \in \Gamma$ for all i). Let*

$$X = A_1A_2 \cap A_4A_5\ , \qquad Y = A_2A_3 \cap A_5A_6\ , \qquad Z = A_3A_4 \cap A_6A_1\ .$$

Then X, Y and Z are collinear.

The line containing these three points is called the *Pascal line* of the hexagon.

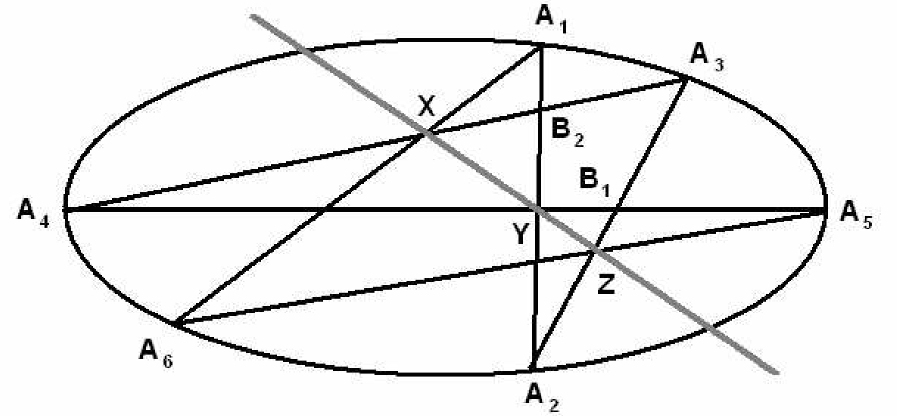


Figure VII.2

We have stated Pascal's Theorem for nonsingular conics, but a version of the result is also true for singular conics given by the union of two lines, provided the hexagon is degenerate in the sense that $\{A_1,\ A_3,\ A_5\}$ lie on one line and $\{A_2,\ A_4,\ A_6\}$ lie on the other. In such a situation, *the conclusion of Pascal's Theorem reduces to the conclusion of Pappus' Theorem*, and hence one can view Pappus' Theorem as a special case of Pascal's Theorem.[8]

SPECIAL CASE. Suppose that Γ in $\mathbb{RP}^2$ is given by the ordinary unit circle and $A_1 \cdots A_6$ is a regular hexagon which is inscribed in Γ. Then it is clear that $A_1A_2||A_41A_5$, $A_2A_3||A_51A_6$ and $A_3A_4||A_61A_1$ (see the illustration below — note that A_1, A_2, A_3, A_4, A_5, A_6 correspond to A, B, C, D, E, F in the drawing), so that X, Y and Z are all ideal points and the Pascal line is equal to the line at infinity.

[7] Blaise Pascal (1623–1662) is known for contributions to a wide range of areas in the mathematical and physical sciences as well as philosophy. Aside from the theorem appearing here, he is particularly recognized for scientific work on fluid mechanics, probability theory, a counting machine which was the prototype for devices like mechanical odometers, as well as the philosophy of science. Most of his philosophical writings were highly religious in nature.

[8] And this is why French and German writers often use phrases translating to "Pascal's Theorem" when referring to the result known as Pappus' (Hexagon) Theorem in the English language.

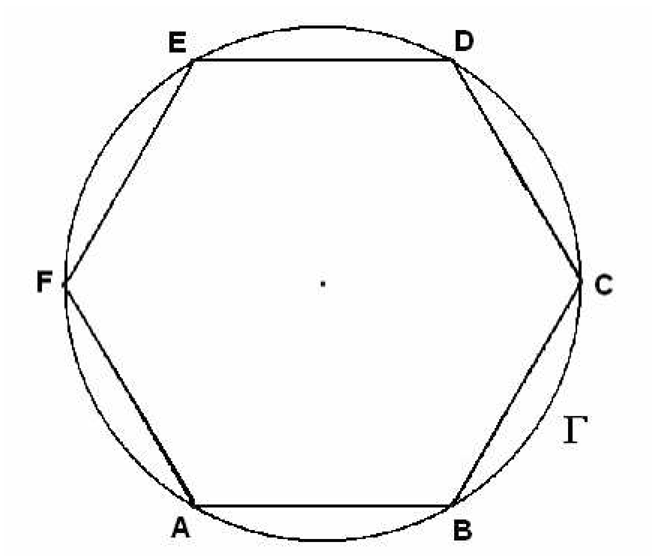


Figure VII.3

Proof of Pascal's Theorem (= Theorem 24). By Steiner's Theorem there is a projective collineation Φ such that $\Phi(A_1) = A_5$ and Φ also has the following properties:

$$\Phi[A_1A_4] = A_5A_4 \qquad \Phi[A_1A_2] = A_5A_2 \qquad \Phi[A_1A_3] = A_5A_3$$

$$\Phi[A_1Z = A_1A_6] = A_5A_6 = A_5Y$$

As suggested by Figure VII.2, we define B_1 to be the point at which A_2A_3 meets A_4A_5, and we define B_2 to be the point at which A_3A_4 meets A_1A_2. Since Φ is a projective collineation, by Exercise VI.3.3 we have the following cross ratio equations:

$$\mathsf{XR}\,(A_1A_4,\ A_1A_3,\ A_1B_2,\ A_1Z) = \mathsf{XR}\,(\,\Phi[A_1A_4],\ \Phi[A_1A_3],\ \Phi[A_1B_2],\ \Phi[A_1Z]\,) =$$
$$\mathsf{XR}\,(A_5B_1,\ A_5A_3,\ A_5A_4,\ A_5Y)$$

By construction, the points Z and B_2 are on A_3A_4, and the points Y and B_1 are on $A - 2A_3$. Therefore Theorem V.14 implies that the first cross ratio in the displayed equation is equal to $\mathsf{XR}(A_4, A_3, B_2, Z)$ and the second is equal to $\mathsf{XR}(B_1, A_3, A_2, Y)$, so that

$$\mathsf{XR}(A_4, A_3, B_2, Z) = \mathsf{XR}(B_1, A_3, A_2, Y)\ .$$

Since $A_4B_1 = A_4A_5$ and $B_2A_2 = A_1A_2$ it follows that $X \in A_4B_1 \cap A_3X \cap B_2A_2$. Thus we also have

$$\mathsf{XR}(B_1, A_3, A_2, Y) = \mathsf{XR}(A_4, A_3, B_2, W)$$

where $W \in A_3A_4 \cap XY$. But the right hand side of the equation is also equal to the cross ratio $\mathsf{XR}(A_4, A_3, B_2, W)$, and therefore $W = Z$ by Theorem V.10. In particular, this implies that

$$Z \in A_3A_4 \cap XY$$

so that X, Y and Z are collinear.■

If we now apply the Augmented Principle of Duality formulated in Section V, we immediately obtain the following result (Brianchon's Theorem),[9] which was originally established without using duality:

[9] Charles Julien Brianchon (1783–1864) worked in mathematics and chemistry; in mathematics he is known for rediscovering Pascal's Theorem and proving the result which bears his name.

THEOREM VII.25. (Brianchon's Theorem) *Suppose that Γ is a nonsingular conic in $\mathbb{FP}^2$ and let the simple hexagon $A_1 \cdots A_6$ be* circumscribed *about Γ (in other words, the lines A_iA_{i+1} are tangent to Gamma for all i, and likewise for A_6A_1). Then the lines A_1A_4, A_2A_3 and A_2A_3 are concurrent.*■

The point of concurrency is called the *Brianchon point*.
SPECIAL CASE. Suppose that Γ in $\mathbb{RP}^2$ is given by the ordinary unit circle and $A_1 \cdots A_6$ is a regular hexagon which is inscribed in Γ. Then the Brianchon point is the center of the circle.

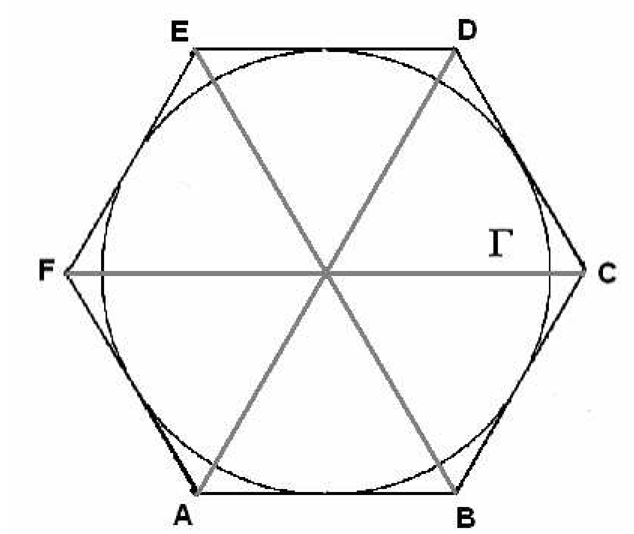


Figure VII.4

There is also a converse to Pascal's Theorem (and hence, by duality, there is also a converse to Brianchon's Theorem).

THEOREM VII.26. *Let $A_1 \cdots A_6$ be a simple hexagon, and let X, Y, Z be defined as in Pascal's Theorem. If these three points are collinear, then there is a conic Γ such that $A_i \in \Gamma$ for all i.*

By Theorem 19, if there is a conic containing the given six points, then it must be nonsingular.

Proof. Let Γ be the unique nonsingular conic containing the first five points $A_1, \cdots, A_5$ and let Φ be the Steiner collineation for Γ with $\Phi(A_1) = A_5$ and $\Phi[A_1A_j] = A_5A_j$ for $j = 2, 3, 4$. It will suffice to prove that $\Phi[A_1A_6] = A_5A_6$. Let B_1 and B_2 be defined as in the proof of Pascal's Theorem. Since $\{A_4, A_3, B_2, Z\}$ and $\{B_1, A_3, A_2, Y\}$ are quadruples of collinear points and

$$X \in A_4B_1 \cap A_3 \cap A_2B_2 \cap YZ$$

Theorem V.14 implies that

$$\mathsf{XR}(A_4, A_3, B_2, Z) = \mathsf{XR}(B_1, A_3, A_2, Y) .$$

This in turn implies the following equation:

$$\mathsf{XR}(A_1A_4, A_1A_3, A_1B_2, A_1Z) = \mathsf{XR}(A_5B_1, A_5A_3, A_5A_2, A_5Y)$$

Since $A_1B_2 = A_1A_2$, $A_1Z = A_1A_6$, $A_5B_1 = A_5A_4$ and $A_5Y = A_5A_6$, the equation above may be rewritten as follows:

$$\mathsf{XR}(A_1A_4, A_1A_3, A_1A_2, A_1A_6) = \mathsf{XR}(A_5A_4, A_5A_3, A_5A_2, A_5A_6)$$

On the other hand, since Φ is a projective collineation, the right hand side is equal to the following:

$$\mathsf{XR}\big(\,\Phi[A_1A_4],\,\Phi[A_1A_3],\,\Phi[A_1A_2],\,\Phi[A_1A_6]\,\big) \;=\; \mathsf{XR}(A_5A_4, A_5A_3, A_5A_2, \Phi[A_1A_6]\,)$$

Therefore it follows that $\Phi[A_1A_6] \;=\; A_5A_6$, which is what we needed to verify in order to complete the proof.$\blacksquare$

The statement of the dual theorem to Theorem 26 is left to the reader (see the exercises).$\blacksquare$

Degenerate cases of Pascal's Theorem

There are analogs of Pascal's Theorem for inscribed simple n-gons where $n \;=\; 3, 4, 5$ (and by duality there are similar analogs of Brianchon's Theorem). Roughly speaking, these are limiting cases in which two consecutive vertices merge into a single point and the line joining the two points converges to the tangent line at the common point. The proofs of these theorems require a simple observation about Steiner collineations.

THEOREM VII.27. *Let Γ be a nonsingular conic, let A and B be points of Γ, and let Φ be a Steiner collineation for Γ such that $\Phi(A) \;=\; B$. If $\mathbf{T}_A$ is the tangent line to Γ at A, then $\Phi[\mathbf{T}_A] \;=\; AB$; if $\mathbf{T}_B$ is the tangent line to Γ at B, then $\Phi[AB] \;=\; \mathbf{T}_B$.*

Proof. Since $B \in \Phi[\mathbf{T}_A]$, we know that $\Phi[\mathbf{T}_A] \;=\; BC$ for some point C. If $D \;\in\; \Phi[\mathbf{T}_A] \cap BC$, then $D \in \Gamma$ by construction. But the only point in $\mathbf{T}_A \cap \Gamma$ is A itself, and therefore we must have $BC \;=\; BA$. Since Φ^{-1} is a Steiner collineation for Γ taking B to X, it follows that $\Phi^{-1}[\mathbf{T}_B] \;=\; AB$, which is equivalent to the desired equation $\Phi[AB] \;=\; \mathbf{T}_B$.$\blacksquare$

Here are the analogs of Pascal's Theorem for inscribed pentagons and quadrilaterals; note that there are two separate analogs for quadrilaterals.

THEOREM VII.28. *Suppose that Γ is a nonsingular conic in $\mathbb{FP}^2$ and let the simple pentagon $A_1 \,\cdots\, A_5$ be inscribed in Γ. Let*

$$X \;=\; A_1A_2 \,\cap\, A_4A_5 \;, \qquad Y \;=\; A_2A_3 \,\cap\, A_5A_1 \;, \qquad Z \;=\; A_3A_4 \,\cap\, \mathbf{T}_{A_1} \;.$$

Then X, Y and Z are collinear.

THEOREM VII.29. *Suppose that Γ is a nonsingular conic in $\mathbb{FP}^2$ and let the simple quadrilateral $A_1 \,\cdots\, A_4$ be inscribed in Γ. Let*

$$X \;=\; \mathbf{T}_{A_1} \,\cap\, A_2A_4 \;, \qquad Y \;=\; A_1A_2 \,\cap\, A_3A_4 \;, \qquad Z \;=\; \mathbf{T}_{A_2} \,\cap\, A_1A_4 \;.$$

Then X, Y and Z are collinear.

THEOREM VII.30. *Suppose that Γ is a nonsingular conic in $\mathbb{FP}^2$ and let the simple quadrilateral $A_1 \,\cdots\, A_4$ be inscribed in Γ. Let*

$$D \;=\; A_1A_3 \,\cap\, A_2A_4 \;, \qquad E \;=\; A_1A_4 \,\cap\, A_2A_3 \;, \qquad F \;=\; \mathbf{T}_{A_1} \,\cap\, \mathbf{T}_{A_2} \;.$$

Then D, E and F are collinear.

The proofs of these theorems are easy variants of the proofs of Pascal's Theorem and are left to the reader as exercises.■

Similarly, formulations and proofs of the duals to all these results are left to the reader as exercises.■

The final degenerate case of Pascal's Theorem requires a special argument. As noted in Appendix A, the cross product of vectors in $\mathbb{F}^3$ satisfies the following condition known as the *Jacobi Identity* (see Theorem A.21):

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \ + \ \mathbf{b} \times (\mathbf{c} \times \mathbf{a}) \ + \ \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) \ = \ \mathbf{0} \ .$$

THEOREM VII.31. *Let ABC be a (projective) triangle inscribed in the nonsingular conic Γ. Let*

$$D \ = \ \mathbf{T}_A \cap BC \ , \qquad E \ = \ \mathbf{T}_B \cap AC \ , \qquad F \ = \ \mathbf{T}_C \cap AB \ .$$

Then D, E and F are collinear.

Proof. By Theorems 5, 10 and 14, the nonsingular conic Γ is projectively equivalent to one defined by an equation of the form $ax^2 + by^2 + cz^2 \ = \ 0$ where none of the coefficients a, b, c is equal to zero. Dividing these by a suitable constant, we may assume $c = -1$. Therefore it suffices to prove the result for conics defined by equations of the form $ax^2 + by^2 - z^2 = 0$.

Let ρ be the 1×3 matrix $(0 \ 0 \ -2)$. If $X \in \Gamma$ and has homogeneous coordinates

$$\xi \ = \ {}^{\mathbf{T}}(x_1 \quad x_2 \quad x_3)$$

then homogeneous coordinates for the tangent line $\mathbf{T}_X$ to Γ at X are given by $\xi^{\#} \ = \ {}^{\mathbf{T}}\xi \ + \ x_3\,\rho$.

Let α, β, γ denote homogeneous coordinates for A, B, C, and let $\alpha^{\#}$, $\beta^{\#}$, $\gamma^{\#}$ denote corresponding homogeneous coordinates for the tangent lines $\mathbf{T}_A$, $\mathbf{T}_B$ and $\mathbf{T}_C$. It will suffice to show that the vectors

$$\alpha^{\#} \ \times \ (\beta \times \gamma) \ , \qquad \beta^{\#} \ \times \ (\gamma \times \alpha) \ , \qquad \gamma^{\#} \ \times \ (\alpha \times \beta)$$

are linearly dependent. However, their sum is equal to

$$[\alpha \ \times \ (\beta \times \gamma) \ + \ \beta \ \times \ (\gamma \times \alpha) \ + \ \gamma \ \times \ (\alpha \times \beta)] \ +$$
$${}^{\mathbf{T}}\rho \ \times \ \big(a_3\,\beta \times \gamma \ + \ b_3\,\gamma \times \alpha \ + \ c_3\,\alpha \times \beta \big)$$

and we claim that this sum vanishes. The term in square brackets vanishes by the Jacobi Identity; to analyze the remaining term(s), we may use the "back–cab formula"

$${}^{\mathbf{T}}\rho \ \times \ (\eta \times \zeta) \ = \ (\rho \cdot \zeta)\eta \ - \ (\rho \cdot \eta)\zeta \ = \ 2(z_3\,\eta \ - \ y_3\zeta)$$

to see that the expression

$${}^{\mathbf{T}}\rho \ \times \ \big(a_3\,\beta \times \gamma \ + \ b_3\,\gamma \times \alpha \ + \ c_3\,\alpha \times \beta \big)$$

is a sum of six terms that cancel each other in pairs.■

As before, the formulation of the dual theorem is left to the reader as an exercise.■

EXERCISES

1. Prove that the conclusion of Theorem 21 is still valid if Γ is completely singular, provided it contains at least two points. [*Hint:* The set of singular points is a geometrical subspace.]

2. Find the equations of the conics in $\mathbb{RP}^2$ which pass through the following five points: with the following homogeneous coordinates:

(i) The five points with the following homogeneous coordinates:

$$\begin{pmatrix}1\\1\\1\end{pmatrix} \qquad \begin{pmatrix}1\\-1\\1\end{pmatrix} \qquad \begin{pmatrix}1\\-1\\-1\end{pmatrix} \qquad \begin{pmatrix}1\\1\\-1\end{pmatrix} \qquad \begin{pmatrix}2\\0\\2\end{pmatrix}$$

(ii) The five points with the following homogeneous coordinates:

$$\begin{pmatrix}1\\1\\0\end{pmatrix} \qquad \begin{pmatrix}1\\0\\1\end{pmatrix} \qquad \begin{pmatrix}0\\1\\1\end{pmatrix} \qquad \begin{pmatrix}1\\2\\2\end{pmatrix} \qquad \begin{pmatrix}1\\2\\3\end{pmatrix}$$

3. Let Φ be the collineation of $\mathbb{RP}^2$ sending the point with homogeneous coordinates ${}^{\mathbf{T}}(1 \quad 1 \quad 2)$ to ${}^{\mathbf{T}}(2 \quad 2 \quad 1)$, and the lines with homogeneous coordinates

$$(2 \quad 0 \quad -1) \qquad (1 \quad -1 \quad 0) \qquad (1 \quad 1 \quad -1)$$

to the lines with homogeneous coordinates

$$(1 \quad 0 \quad -2) \qquad (1 \quad 1 \quad -4) \qquad (1 \quad 2 \quad -6)$$

respectively. Find the equation of the conic Γ for which Φ is a Steiner collineation (using Theorem 22).

4. Let Γ be the ellipse defined by the affine equation $x^2 + 3y^2 = 4$ in $\mathbb{R}^2$ (hence its projectivization has no ideal points). If T is the unique projective collineation of $\mathbb{RP}^2$ sending $\mathrm{J}(\pm 2, 0)$ and $\mathrm{J}(0, -\frac{2}{3}\sqrt{3})$ to themselves, and sending $\mathrm{J}(-1,1)$ to $\mathrm{J}(1,1)$, then T is a Steiner collineation for Γ. Likewise, if S is the unique projective collineation of $\mathbb{RP}^2$ sending $\mathrm{J}(\pm 2, 0)$ and $\mathrm{J}(0, \frac{2}{3}\sqrt{3})$ to themselves, and sending $\mathrm{J}(-1,1)$ to $\mathrm{J}(1,1)$, then S is also a Steiner collineation for Γ. Show that S and T must be distinct projective collineations. [*Hint:* If $S = T$, then this map fixes the four points on Γ where it meets the x- and y-axes. What does the Fundamental Theorem of Projective Geometry imply about $S = T$ in this case?]

5. State the duals of Theorems 19, 21 and 27–31 (the duality principle implies that these dual results are automatically valid).

6. Prove Theorems 27–30 and their converses.

7. Let Γ be a nonsingular conic in $\mathbb{FP}^2$, and let $\{A, B, C\}$ and $\{A', B', C'\}$ be two disjoint noncollinear sets of points on Γ. Prove that the lines AA', BB' and CC' are concurrent. — A drawing and hints appear on the next page.

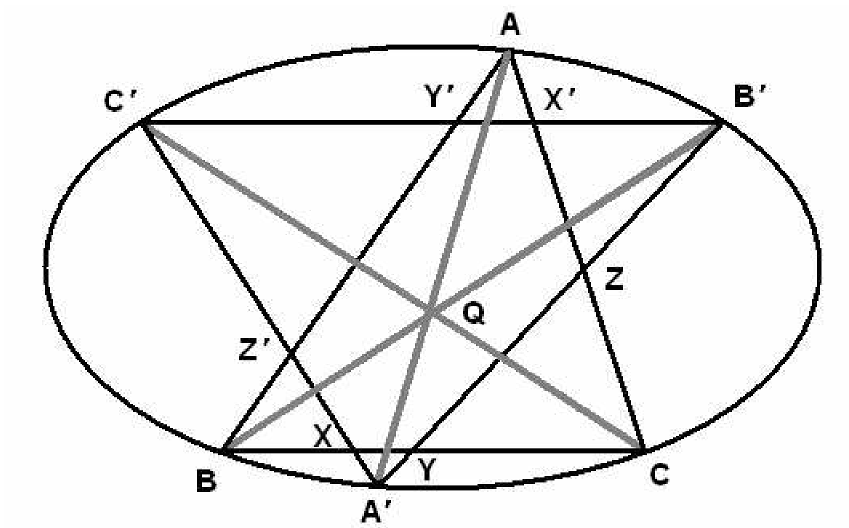


Figure VII.5

[*Hint:* Define X, Y, Z and X', Y', Z' as suggested in the figure. First prove that AA', XX' and YY' are concurrent using Pascal's Theorem. Clearly corresponding results hold for BB' and CC'. Let $U \in C'Y \cap BX'$, and prove that Z', U and Q are collinear by Pappus' Theorem. Also show that ZU, BB' and CC' are concurrent using Pappus' Theorem for $\{C', X', B\}$ and $\{B, Y, C\}$. Finally, apply Pascal's Theorem to $AB'C'A'BC$ to show that BB', CC' and ZZ' are concurrent. Using similar results for XX' and YY' and the previous concurrency relations involving AA', BB' and CC', prove that all six lines are concurrent.]

What is the dual of the preceding result?

8. Let Γ, $\{A, B, C\}$ and $\{A', B', C'\}$ be as in Exercise 6. Prove that the six lines determined by the triangles ABC and $A'B'C'$ form the sides of a hexagon that is tangent to another conic. [*Hint:* The hexagon is $Y'X'ZYXZ'$. Apply the converse to Brianchon's Theorem.]

9. Show that a set of six points on a nonsingular conic Γ determines sixty simple hexagons (in general these sixty hexagons have distinct Pascal lines, and the footnote on page 152 of Fishback contains further information on the totality of all such configurations).