Finite Geometry

Chris Godsil
Combinatorics & Optimization
University of Waterloo

Preface

We treat some topics in finite geometry.

Contents

Chapter 1

Examples

We introduce projective and affine spaces over fields.

1.1 Projective Space and Subspaces

Let $\mathbb{F}$ be a field and let V be the vector space $\mathbb{F}^{n\times 1}$. The *projective space of rank d* consists of the subspaces of V. The 1-dimensional subspaces of V are the *points* and the 2-dimensional subspaces are the lines. The subspaces of V of dimension $n-1$ are called *hyperplanes*.

We can represent each point by a non-zero element x of V, provided we understand that any non-zero scalar multiple of x represents the same point. We can represent a subspace of V with dimension k by an $n\times k$ matrix M over $\mathbb{F}$ with linearly independent columns. The column space of M is the subspace it represents. Clearly two matrices M and N represent the same subspace if and only if there is an invertible $k\times k$ matrix A such that $M = NA$. (The subspace will be determined uniquely by the reduced column-echelon form of M.)

If M represents a hyperplane, then $\dim(\ker M^T) = 1$ and so we can specify the hyperplane by a non-zero element a of $\mathbb{F}^{n\times 1}$ such that $a^T M = 0$. Then x is a vector representing a point on this hyperplane if and only if $a^T x = 0$.

We introduce the *Gaussian binomial coefficients*. Let q be fixed and not equal to 1. We define

$$[n] := \frac{q^n - 1}{q-1}.$$

If the value of q needs to be indicated we might write $[n]_q$. We next define $[n]!$ by declaring that $[0] := 1$ and

$$[n+1]! = [n+1][n]!.$$

Note that $[n]$ is a polynomial in q of degree $n-1$ and $[n]!$ is a polynomial in q of degree $\binom{n}{2}$. Finally we define the Gaussian binomial coefficient by

$$\begin{bmatrix} n \\ k \end{bmatrix} := \frac{[n]!}{[k]![n-k]!}.$$

1.1.1 Theorem. *Let V be a vector space of dimension n over a field of finite order q. Then the number of subspaces of V with dimension k is $\left[{n \atop k}\right]$.*

Proof. First we count the number N_r of $n \times r$ matrices over $GF(q)$ with rank r. There are $q^n - 1$ non-zero vectors in V, so $N_1 = q^n - 1$.

Suppose A is an $n \times r$ matrix with rank r. Then there are $q^r - 1$ non-zero vectors in $\operatorname{col}(A)$, and therefore there are $q^n - q^r$ non-zero vectors not in $\operatorname{col}(A)$. If x is one of these, then (A, x) is an $n \times (r+1)$ matrix with rank r, and therefore

$$N_{r+1} = (q^n - q^r)N_r.$$

Hence

$$N_r = (q^n - q^{r-1}) \cdots (q^n - 1) = q^{\binom{r}{2}}(q-1)^r \frac{[n]!}{[n-r]!}.$$

Note that N_n is the number of invertible $n \times n$ matrices over $GF(q)$. Count pairs consisting of a subspace U of dimension r and an $n \times r$ matrix A such that $U = \operatorname{col}(A)$. If ν_r denotes the number of r-subspaces then

$$N_r = \nu_r q^{\binom{r}{2}}(q-1)^r[r]!.$$

This yields the theorem. □

Suppose U_1 and U_2 are subspaces of V. We say that U_1 and U_2 are *skew* if $U_1 \cap U_2 = \{0\}$; geometrically this means they are skew if they have no points in common. We say that U_1 and U_2 are *complements* if they are skew and $V = U_1 + U_2$; in this case

$$\dim V = \dim U_1 + \dim U_2.$$

Now suppose that U and W are complements in V and $\dim(U) = k$. If H is a subspace of V that contains U, define $\rho(H)$ by

$$\rho(H) = H \cap W.$$

We claim that ρ is a bijection from the set of subspaces of V that contain U and have dimension $k + \ell$ to the subspaces of W with dimension ℓ.

We have $H + W = V$ and therefore

$$\begin{aligned} n = \dim(H + W) &= \dim(H) + \dim(W) - \dim(H \cap W) \\ &= k + \ell + n - k - \dim(H \cap W) \\ &= n + \ell - \dim(H \cap W). \end{aligned}$$

This implies that $\dim(H \cap W) = \ell$. It remains for us to show that ρ is a bijection.

If W_1 is a subspace of W with dimension ℓ, then $U + W_1$ is a subspace of V with dimension $k + \ell$ that contains U. Then

$$\rho(U + W_1) = W_1,$$

which shows that ρ is surjective. Suppose $\rho(H_1) = \rho(H_2)$. Then

$$H_1 \cap W = H_2 \cap W$$

and so both H_1 and H_2 contain $U + (H_1 \cap W)$. Since these three spaces all have dimension $k + \ell$, it follows that they are equal. Therefore ρ is injective.

We also notes that H and K are subspaces of V that contain U and $H \leq K$, then $\rho(H) \leq \rho(K)$. Therefore ρ is an inclusion-preserving bijection from the subspaces of V that contain U to the subspaces of W. The subspaces of W form a projective space of rank $n - \dim(U)$ and so it follows that we view the subspaces of V that contain U as a projective space.

1.1.2 Lemma. *Let V be a vector space of dimension n over a field of order q, and let U be a subspace of dimension k. The number of subspaces of V with dimension ℓ that are skew to U is $q^{k\ell}\left[{n-k \atop \ell}\right]$.*

Proof. The number of subspaces of V with dimension $k + \ell$ that contain U is $\left[{n-k \atop \ell}\right]$. If W_1 has dimension ℓ and is skew to U, then $U + W_1$ is a subspace of dimension $k + \ell$ that contains U. Hence the subspaces of dimension $k + \ell$ that contain U partition the set of subspaces of dimension ℓ that are skew to U. The number of subspaces of dimension $k + \ell$ in V that contain U is $\left[{n-k \atop \ell}\right]$. We determine the number of complements to U in a space W of dimension m that contains U.

We identify W with $\mathbb{F}^{m\times 1}$. Since $\dim W = m$ and $\dim U = k$, we may assume that U is the column space of the $m \times k$ matrix

$$\begin{pmatrix} I_k \\ 0 \end{pmatrix}$$

Suppose W_1 is a subspace of W with dimension $m - k$. We may assume that W is the column space of the $m \times (m - k)$ matrix

$$\begin{pmatrix} A \\ B \end{pmatrix}$$

where B is $(m - k) \times (m - k)$. Then W_1 is a complement to U if and only if the matrix

$$\begin{pmatrix} I_k & A \\ 0 & B \end{pmatrix}$$

is invertible, and this hold if and only if B is invertible. If B is invertible, then W_1 is the column space of

$$\begin{pmatrix} AB^{-1} \\ I_{m-k} \end{pmatrix}.$$

So there is a bijection from the set of complements to U in W to the set of $m \times (m - k)$ matrices over $\mathbb{F}$ of the form

$$\begin{pmatrix} M \\ I \end{pmatrix},$$

and therefore the number of complements of U is equal to $q^{k(m-k)}$, the number of $k \times (m - k)$ matrices over $\mathbb{F}$. □

1.2 Affine Spaces

We define *affine n-space* over $\mathbb{F}$ to be $\mathbb{F}^n$, equipped with the relation of affine dependence. A sequence of points $v_1, \dots, v_k$ from $\mathbb{F}^n$ is *affinely dependent* if there are scalars $a_1, \dots, a_k$ not all zero such that

$$\sum_i a_i = 0, \quad \sum_i a_i v_i = 0.$$

We also say that v is an *affine linear combination* of $v_1, \dots, v_k$ if

$$v = \sum_i a_i v_i$$

where

$$\sum_i a_i = 1.$$

Thus if v is an affine linear combination of $v_1, \dots, v_k$, then the vectors $-v, v_1, \dots, v_k$ are affinely dependent.

Note that if $v \neq 0$ and $a \neq 1$ then the vectors v, av are not affinely dependent. In particular if $v \neq 0$, then $0, v$ is not affinely dependent. In affine spaces the zero vector does not play a special role.

If u and v are distinct vectors, then the set

$$\{au + (1-a)v : a \in \mathbb{F}\}$$

consist of all affine linear combinations of u and v. If $\mathbb{F} = \mathbb{R}$ then it is the set of points on the straight line through u and v; in any case we call it the affine line through u and v. A subset S of V is an *affine subspace* if it is closed under taking affine linear combination of its elements. Equivalently, S is a subspace if, whenever it contains distinct points u and v, it contains the affine line through u and v. (Prove it.) A single vector is an affine subspace. The affine subspaces $\mathbb{F}^n$ are the cosets of the linear subspaces.

Suppose $\mathcal{A}$ denotes the elements of $\mathbb{F}^{n+1}$ with last coordinate equal to 1. Then a subset of S of $\mathcal{A}$ is linearly dependent in $\mathbb{F}^{n+1}$ if and only if it affinely dependent. This allows us to identify affine n-space over $\mathbb{F}$ with a subset of projective n-space over $\mathbb{F}$. (In fact projective n-space is the union of $n+1$ copies of affine n-space.)

1.3 Coordinates

We start with the easy case. If $\mathcal{A}$ is the affine space $\mathbb{F}^n$, then each point of $\mathcal{A}$ is a vector and the coordinates of a point are the coordinates of the associated vector.

Now suppose $\mathcal{P}$ is the projective space associated to $\mathbb{F}^n$. Two non-zero vectors x and y represent the same point if and only if there is a non-zero scalar a such that $y = ax$. Thus a point is an equivalence class of non-zero vectors.

As usual it is often convenient to represent an equivalence class by one of its elements. Here there is no canonical choice, but we could take the representative to be the vector with first non-zero coordinate equal to 1. Normally we will **not** do this.

The map that takes a vector in $\mathbb{F}^n$ to its i-th coordinate is called a *coordinate function*. It is an element of the dual space of $\mathbb{F}^n$. The sum of a set of coordinate functions is a function on $\mathbb{F}^n$. If $f_1, \dots, f_k$ is a set of coordinate functions then the product $f_1 \cdots f_k$ is a function on $\mathbb{F}^n$. The set of all linear combinations of products of coordinate functions is the algebra of polynomials on $\mathbb{F}^n$. Many interesting structures can be defined as the set of common zeros of a collection of polynomials.

Defining functions on projective space is trickier, because each point is represented by a set of vectors. However if p is a homogeneous polynomial in n variables with degree k and $x \in \mathbb{F}^n$, then

$$p(ax) = a^k p(x).$$

Thus it makes sense to consider structures defined as the set of common zeros of a set of homogeneous polynomials.

If we are working over the reals, another approach is possible. If x is a unit vector in $\mathbb{R}^n$, then the $n \times n$ matrix xx^T represents orthogonal projection onto the 1-dimensional subspace spanned by x. Thus we obtain a bijection between the points of the projective space and the set of symmetric $n \times n$ matrices X with $\operatorname{rk}(X) = 1$ and $\operatorname{tr}(X) = 1$. However it is a little tricky to decide if three such matrices represent collinear points. (A similar trick works for complex projective space; we use matrices xx^*, which are Hermitian matrices with rank one.)

Chapter 2

Projective and Affine Spaces

We start by considering geometries in the abstract, and then projective and affine geometries in particular.

2.1 Lots of Definitions

An *incidence structure* $\mathcal{I}$ consists of a set of points P, a set of blocks L and an incidence relation between the points and blocks. If the point p is incident with the block ℓ then we say p is on ℓ, and write $p \in \ell$. A *linear space* is an incidence structure with the property that any pair of distinct points lies in a unique block, and any block contains at least two points. In this case blocks are usually called lines. Any two lines in a linear space have at most one point in common. The unique line through the points p and q will be denoted by $p \vee q$. A set of points is *collinear* if it is contained in some line.

A subspace of a linear space is a subset S of its points with the property that if $p \in S$ and $q \in S$ then $p \vee q \subseteq S$. (The last is an abuse of notation, and is intended to indicate that all points in $p \vee q$ lie in S.) We can make S into a linear space by defining its line set to be the lines of $\mathcal{I}$ which meet it in at least two points. The intersection of any two subspaces is a subspace. The empty set and the entire space are subspaces. The *join* of two subspaces H and K is defined to be the intersection of all subspaces which contain both H and K, and is denoted by $H \vee K$. Every subset S of the points of a linear space determines a subspace, namely the intersection of all subspaces which contain it. This subspace is said to be *spanned* by S.

A *rank function* rk on a set P is a function from the subsets of P to the non-negative integers such that:

(a) if $A \subseteq P$ then $0 \leq \mathrm{rk}(A) \leq |A|$ and

(b) if $B \subseteq A$ then $\mathrm{rk}(B) \leq \mathrm{rk}(A)$.

If, in addition

1. $\mathrm{rk}(A \cup B) + \mathrm{rk}(A \cap B) \leq \mathrm{rk}(A) + \mathrm{rk}(B)$

then we say the rank function is submodular. A set equipped with a submodular rank function is called a *matroid*. A *flat* in a matroid is a subset F such that, if $p \notin F$ then $\mathrm{rk}(p \cup F) > \mathrm{rk}(F)$. A *combinatorial geometry* is a set P, together with a submodular rank function rk such that if $A \subseteq P$ and $|A| \leq 2$ then $\mathrm{rk}(A) = |A|$. Every combinatorial geometry can be regarded as a linear space with the flats of rank one as its points and the flats of rank two as its lines.

We can often make a linear space into a matroid as follows. A set of distinct subspaces $S_0, \dots, S_r$ of a linear space $\mathcal{L}$ such that

$$S_0 \subset \cdots \subset S_r$$

is called a *flag*. Define the rank $\mathrm{rk}(A)$ of a subspace A to be the maximum number of non-empty subspaces in a flag consisting of subspaces of A. We then define the rank of a subset to be the rank of the subspace spanned by it. If we refer to a rank function on a linear space without otherwise specifying it, this is the function we will mean. This function trivially satisfies conditions (a) and (b) above, but may not be submodular. When it is, we say that the lattice of subspaces of $\mathcal{L}$ is *semimodular*. The lattice of subspaces of a combinatorial geometry, viewed as a linear space, is always semimodular. If (c) holds with equality for subspaces then the subspace lattice is *modular*. The lattice of subspaces of a vector space provide the most important example of this. It is left as an exercise to show that the maximal proper flats of a combinatorial geometry all have the same rank. These flats are called the *hyperplanes* of the geometry. The flats of rank two are its *lines* and the flats of rank three are its *planes*. The rank of a combinatorial geometry is the maximum value of its rank function. We will always assume this is finite, even if the point set is not.

A *collineation* of a linear space is bijection of its point set onto itself which maps each line onto a line. Similarly we define collineations between distinct linear spaces. It should be clear that the image of a subspace under a collineation is a subspace. The set of all collineations of a linear space onto itself is its *collineation group*. Two linear spaces are isomorphic if there is a collineation from one onto the other.

2.2 Axiomatics

A *projective geometry* is officially a linear space such that

(a) if x, y and z are non-collinear points and the line ℓ meets $x \vee y$ and $x \vee z$ in distinct points then it meets $y \vee z$,

(b) every line contains at least three points.

The first condition is known as *Pasch's axiom*. Linear spaces satisfying the second condition are often said to be *thick*. We show that $PG(n, \mathbb{F})$ satisfies these axioms. Suppose that x, y, z are three non-collinear points and that ℓ is a line meeting $x \vee y$ and $x \vee z$ in distinct points. Then

$$\operatorname{rk}(\ell \cap (y \vee z)) = \operatorname{rk}(\ell) + \operatorname{rk}(y \vee z) - \operatorname{rk}(\ell \vee (y \vee z)). \tag{2.1}$$

Our conditions imply that

$$\ell = (\ell \cap (x \vee y)) \vee (\ell \cap (x \vee z)).$$

Since ℓ thus contains two points of the subspace $x \vee y \vee z$, it must be contained in it. It follows that $\ell \vee (y \vee z))$ is also contained in it. Now x, y and z are not collinear and therefore $(x \vee y) \cap z = \emptyset$. Thus

$$\operatorname{rk}(x \vee y \vee z) = \operatorname{rk}(x \vee y) + \operatorname{rk}(z) - \operatorname{rk}((x \vee y) \cap z) = \operatorname{rk}(x \vee y) + \operatorname{rk}(z) = 3.$$

and consequently $\operatorname{rk}(\ell \vee (y \vee z)) \leq 3$. From (2.1) we now infer that $\operatorname{rk}(\ell \cap (y \vee z)) \geq 1$, which implies that ℓ meets $y \vee z$. Since any field has at least two elements, any line of $PG(n, \mathbb{F})$ contains at least three points. This proves our claim.

We make some comments about projective planes. The standard description of a projective plane is that it is an incidence structure of points and lines such that

(a) any two distinct points lie on a unique line,

(b) any two distinct lines have a unique point in common,

(c) there are four points, such that no three are collinear.

The third axiom is equivalent to the requiring that every line should have at least three points, and that there be at least two lines. (The proof of this claim is an important exercise.) It is easy to verify that any projective plane is a projective geometry of rank three; the converse is less immediate.

The main result of the first part of this course will be that any finite projective geometry with rank n at least four is isomorphic to $PG(n-1, \mathbb{F})$ for some finite field $\mathbb{F}$. (This result also holds for infinite projective geometries of finite dimension, if we allow $\mathbb{F}$ to be non-commutative.) We begin working towards a proof of this.

2.2.1 Lemma. *Let $\mathcal{G}$ be a projective geometry. If H is a subspace of $\mathcal{G}$ and p is a point not on H then $p \vee H$ is the union of the lines through p which contain a point of H.*

Proof. Let S be the set of all points which lie on a line joining p to a point of H. We will show that S is a subspace of $\mathcal{G}$. Suppose that ℓ is a line containing the points x and y from S. By the definition of S, the point y is on line joining p to a point in H and if $x = p$ then this line must be ℓ. If both x and y lie in H then $\ell \in H$, since H is a subspace. Thus we may assume that x and y are both distinct from p and do not lie in H. It follows that both x and y lie on lines through p which meet H. Suppose that they meet H in x' and y' respectively. The line ℓ meets the line $p \vee x'$ and $p \vee y'$ in distinct points; therefore it must intersect $x' \vee y'$ in some point q. If u is a point on ℓ then the line $p \vee u$ meets $y \vee y'$ in p and $y \vee q$ in u. hence it must meet the line $q \vee y'$, which lies in H. As u was chosen arbitrarily on ℓ, it follows that each point of ℓ lies on a line joining p to a point of H. This shows that all points on ℓ lie in S, and so S is a subspace. Any subspace which contains both p and H must contain all points on the lines joining p to points of H. Thus S is the intersection of all subspaces containing p and H, i.e., $S = p \vee H$. □

2.2.2 Corollary. *Let p be a point not in the subspace H. Then each line through p in $p \vee H$ intersects H.*

Proof. Let ℓ be a line through p in $p \vee H$. If x is point other than p in ℓ then x lies on a line through p which meets H. Since x and p lie on exactly one line, it must be ℓ. Thus ℓ meets H. □

We can now prove one of classical results in projective geometry, due to Veblen and Young.

2.2.3 Theorem. *A linear space is a projective geometry if and only if every subspace of rank three is a projective plane.*

Proof. We prove that any two lines in a projective geometry of rank three must intersect. This implies that projective geometries of rank three are projective planes. Suppose that ℓ_1 and ℓ_2 are two lines in a rank three geometry. Let p be a point in ℓ_1 but not in ℓ_2. From the previous corollary, each line through p in $p \vee \ell_2$ must meet ℓ_2. Since $p \vee \ell_2$ has rank at least three, it must be the entire geometry. Hence $\ell_1 \in p \vee \ell_2$ and so it meets ℓ_2 as required. To prove the converse, note that Pasch's axiom is a condition on subspaces of rank three, that is, it holds in a linear space if and only if it holds in all subspaces of rank three. But as we noted earlier, if every two lines in a linear space of rank three meet then it is trivial to verify that Pasch's axiom holds in it. □

2.3 The Rank Function of a Projective Geometry

One of the most important properties of projective geometries is that their rank functions are modular. Proving this is the main goal of this section. A useful by-product of our will be the result that a linear space is a projective geometry if and only if all subspaces with rank three are projective planes. (If there are no projective planes then our geometry has rank at most two, and is thus either a single point or a line.) Note that if p is a point and H a subspace in any linear space then $\operatorname{rk}(p \vee H) \geq \operatorname{rk}(H) + 1$. We will use this fact repeatedly.

2.3.1 Lemma. *Let H and K be two subspaces of a projective geometry such that $H \subset K$ and let p be a point not in K. Then $p \vee H \subset p \vee K$.*

Proof. Clearly $p \vee H \subseteq p \vee K$ and if $p \vee H = p \vee K$ then $K \subseteq p \vee H$. If the latter holds and $k \in K \setminus H$ then the line $p \vee k$ must contain a point, h say, of H. This implies that $p \in h \vee k$ and, since $h \vee k \subseteq K$, that $p \in K$. □

2.3.2 Corollary. *Let H be a subspace of a projective geometry and let p be a point not in H. Then H is a maximal subspace of $p \vee H$.*

Proof. Let K be a subspace of $p \vee H$ strictly containing H. If $p \in K$ then $K = p \vee H$. If $p \notin K$ then, by the previous lemma, $p \vee H$ is strictly contained in $p \vee K$. Since this contradicts our assumption that $K \subseteq p \vee H$, our result is proved. □

2.3.3 Theorem. *All maximal subspaces of a projective geometry have the same rank.*

Proof. We will actually prove a more powerful result. Let H and K be two distinct maximal subspaces. Let h be point in $H \setminus K$ and let k be a point in $K \setminus H$. The line $h \vee k$ cannot contain a second point, h' say, of H since then we would have $k \in h \vee h' \subseteq H$. Similarly $h \vee k$ cannot contain a point of K other than k. By the first axiom for a projective geometry, $h \vee k$ must contain a point p distinct from h and k, and by what we have just shown, $p \notin H \cup K$. Since H and K are maximal $p \vee H = p \vee K$. By Corollary 3.2, each line through p must contain a point of H and a point of K. Using p we construct a mapping ϕ_p from H into K. If $h \in H$ then

$$\phi_p(h) := (p \vee h) \cap K.$$

If $\phi_p(h_1) = \phi_p(h_2)$ then the lines $p \vee h_1$ and $p \vee h_2$ have two points in common, and therefore coincide. This implies that they meet H in the same point and hence ϕ_p is injective. If $k \in K$ then $k \vee p$ must contain a point h' say, of H. We have $\phi_p(h') = k$, whence ϕ_p is surjective. Thus we have shown that ϕ_p is a bijection. We prove next that ϕ_p maps subspaces onto subspaces. Let L be a subspace of H. Then $\phi_p(L)$ lies in $(p \vee L) \cap K$. Conversely, if $x \in (p \vee L) \cap K$ then x is on a line joining p to a point of L and so $x \in \phi_p(L)$. Hence $\phi_p(L) = (p \vee L) \cap K$. Since $p \vee L$ is a subspace, so is $(p \vee L) \cap K$. As ϕ_p is bijective on points, it must map distinct subspaces of H onto distinct subspaces of K. A similar argument to the above shows that ϕ_p^{-1} maps subspaces of K onto subspaces of H. Consequently we have shown that ϕ_p induces an isomorphism from the lattice of subspaces of H onto the subspaces of K. This implies immediately that H and K have the same rank. (It is also worth noting that it implies that ϕ_p is a collineation—it must map subspaces of rank two to subspaces of rank two.) □

A more general form of the next result is stated in the Exercises.

2.3.4 Lemma. *Let H and K be subspaces of a projective geometry and let p be a point in H. Then $(p \vee K) \cap H = p \vee (H \cap K)$.*

Proof. As $H \cap K$ is contained in both $p \vee K$ and H and as $p \in H$, it folllows that $p \vee (H \cap K) \subseteq (p \vee K) \cap H$. Let x be a point in $(p \vee K) \cap H$. By Corollary 3.2, there is a point k in K such that $x \in p \vee k$. Now $p \vee k = p \vee x$ and so $k \in p \vee x$. Since $x \in H$ then this implies that $p \vee x \subseteq H$ and thus that k lies in H as well as K. Summing up, we have shown that if $x \in (p \vee K) \cap H$ then $x \in p \vee k$, where $k \in H \cap K$, i.e., that $x \in p \vee (H \cap K)$. □

2.3.5 Theorem. *If H and K are subspaces of a projective geometry then*

$$\mathrm{rk}(H \vee K) + \mathrm{rk}(H \cap K) = \mathrm{rk}(H) + \mathrm{rk}(K).$$

Proof. We use induction on $\mathrm{rk}(H) - \mathrm{rk}(H \cap K)$. Suppose first that this difference is equal to one. This implies $H \cap K$ is maximal in H. From the previous lemma we now deduce that

$$\mathrm{rk}(H) - \mathrm{rk}(H \cap K) = 1. \tag{2.2}$$

If $p \in H \backslash K$ then, using the maximality of $H \cap K$ in H, we find that $p \vee (H \cap K) = H$ and $H \vee K = p \vee K$. By Corollary 4.2, it follows that K is maximal in $H \vee K$ and so

$$\mathrm{rk}(H \vee K) - \mathrm{rk}(K) = 1. \tag{2.3}$$

Subtracting (2.2) from (2.3) and rearranging yields the conclusion of the Theorem. Assume now that $H \cap K$ is not maximal in H. Then we can find a point $p \in H \cap K$ such that $p \vee (H \cap K) \neq H$. Suppose $L = p \vee (H \cap K)$. Then $H \cap K$ is maximal in L (by Corollary 4.2) and so, by what we have already proved,

$$\mathrm{rk}(L \vee K) + \mathrm{rk}(L \cap K) = \mathrm{rk}(L) + \mathrm{rk}(K). \tag{2.4}$$

Next we note that $\mathrm{rk}(H) - \mathrm{rk}(L \vee K) < \mathrm{rk}(H) - \mathrm{rk}(H \cap K)$ and so by induction we have

$$\mathrm{rk}(H \vee (L \vee K)) + \mathrm{rk}(H \cap (L \vee K)) = \mathrm{rk}(L \vee K) + \mathrm{rk}(H). \tag{2.5}$$

Now $L \vee K = p \vee (H \cap K) \vee K = p \vee K$. By the previous lemma then,

$$H \cap (L \vee K) = H \cap (p \vee K) = p \vee (H \cap K) = L.$$

Furthermore $H \vee (L \vee K) = H \vee K$, and so (2.5) can be rewritten as

$$\mathrm{rk}(H \vee K) + \mathrm{rk}(L) = \mathrm{rk}(L \vee K) + \mathrm{rk}(H). \tag{2.6}$$

Since $L \cap K = H \cap K$, we can now derive the theorem by adding (2.4) to (2.6) and rearranging. □

An important consequence of this theorem is that that the rank of a subspace of a projective geometry spanned by a set S is at at most $|S|$. In particular, three pairwise non-collinear points must span a plane, rather some subspace of larger rank.

2.4 Duality

Let H and K be two maximal subspaces of a projective geometry with rank n. Then $\mathrm{rk}(H \vee K) = n$ and from Theorem 4.5 we have

$$\mathrm{rk}(H \cap K) = \mathrm{rk}(H) + \mathrm{rk}(K) - \mathrm{rk}(H \vee K) = (n-1) + (n-1) - n = n - 2.$$

Thus any pair of maximal subspaces intersect in a subspace of rank $n-2$, and therefore we can view the subspaces of rank $n-1$ and the subspaces of rank $n-2$ as the points and lines of a linear space. We call this the *dual* of our projective geometry. (Linear spaces in general do not have duals.)

2.4.1 Theorem. *The dual of a projective geometry is a projective geometry.*

Proof. We first show that each line in the dual lies on at least three points. Let K be space of rank $n-2$ and let H_1 be a hyperplane which contains it. Since H_1 is not the whole space, there must be point p not in it. Then K is maximal in $p \vee K$ and so $p \vee K$ is a subspace of rank $n-1$ on k. It is not equal to H, because p is in it. Now choose a point q in $H \setminus K$. The line $p \vee q$ must contain a third point, x say. If $x \in H$ then $p \in x \vee q \subseteq H$, a contradiction. Similarly x cannot lie in K and so it follows that $x \vee K$ is a third subspace of rank $n-1$ on K. (We also used this argument in the proof of Theorem 4.3.) Now we should verify the second axiom. However we will show that any two subspaces of rank $n-2$ intersecting in a subspace of rank $n-3$ lie in a subspace of rank $n-1$. This implies that any two lines in the dual which line a subspace of rank three must intersect, and so all rank three subspaces are projective planes. An appeal to Corollary 4.6 now completes the proof. So, suppose that K_1 and K_2 are subspaces with rank $n-2$ which meet in a subspace of rank $n-3$. Then

$$\mathrm{rk}(K_1 \vee K_2) = \mathrm{rk}(K_1) + \mathrm{rk}(K_2) - \mathrm{rk}(K_1 \cap K_2) = (n-2) + (n-2) - (n-3) = n-1$$

and $K_1 \vee K_2$ has rank $n-1$ as required. □

Our next task is to determine the relation between the subspaces of a projective geometry and those of its dual. It is actually quite simple—it is equality.

2.4.2 Lemma. *Let $\mathcal{G}$ be a projective geometry and let L be a subspace of it. Then the hyperplanes which contain L are a subspace in the dual of $\mathcal{G}$.*

Proof. Suppose that $\mathcal{G}$ has rank n. The lines of the dual are the sets of hyperplanes which contain a given subspace of rank $n-2$. Suppose that if K is a subspace of rank $n-2$ and H_1 and H_2 are two maximal subspaces which contain K. If both H_1 and H_2 contain L then $L \subseteq H_1 \cap H_2 = K$. This proves the lemma. □

It can be shown that, if the lattice of subspaces of $\mathcal{G}$ is semimodular, any subspace is the intersection of the hyperplanes which contain it. As we have no immediate use for this, we have assigned it as an exercise, but it is worth noting that it inplies that each subspace of a projective geometry is the intersection of the hyperplanes which contain it.

It is clear from the axioms that any subspace H of a projective geometry is itself a projective geometry. The previous lemma yields that the hyperplanes which contain H are also the points of a projective geometry. Furthermore, if K is a subspace of rank m contained in H then the maximal subspaces of H which contain K are again the points of a projective geometry. Applying duality to this last remark, we see that the subspaces of rank $m+1$ in H which contain K are the points of projective geometry. We will denote this geometry by H/K, and refer to it as an *interval* of the original geometry. Duality is a useful, but somewhat slippery concept. It will reappear in later sections, sometimes saving half our work.

2.5 Affine Geometries

We have already met the affine spaces $AG(n, \mathbb{F})$. An *affine geometry* is defined as follows. Let $\mathcal{G}$ be a projective geometry and let H be a hyperplane in it. If S is set of points in $\mathcal{G} \setminus H$, define $\mathrm{rk}_H(S)$ to be $\mathrm{rk}(S)$. This can be shown to be a submodular rank function on the points not on H, and the combinatorial geometry which results is an *affine geometry*. (It will sometimes be denoted by $\mathcal{G}^H$.) From Lemma 2.2 we see that $AG(n, \mathbb{F})$ can be obtained from $PG(n, \mathbb{F})$ in this way. The flats of $\mathcal{A}$ are defined to be the subsets of the form $K \setminus H$, where K is a subspace of $\mathcal{G}$. They will be referred to as *affine subspaces*; these are all linear subspaces. However, in some cases there will be linear subspaces which are not flats. (This point will be considered in more detail later in this section.) If K_1 and K_2 are two subspaces of $\mathcal{G}$ such that $K_1 \cap K_2 \subseteq H$ and

$$\mathrm{rk}(K_1 \cap K_2) = \mathrm{rk}(K_1) + \mathrm{rk}(K_2) - \mathrm{rk}(K_1 \vee K_2),$$

we say that they are *parallel*. The most important cases are parallel hyperplanes and parallel lines. The hyperplane H is often called the “hyperplane at infinity”, since it is where parallel lines meet. From the definition we see that two disjoint subspaces of an affine geometry are parallel if and only if the dimension of their join is ‘as small as possible’. In particular, two lines are parallel if and only they are disjoint and coplanar. It is not too hard to verify that parallelism is an equivalence relation on the subspaces of an affine geometry. (This is left as an exercise.) The lines of $\mathcal{G}$ which pass through a given point of H partition the point set of the affine geometry. We call such a set of lines a *parallel class*. Any set of parallel lines can be extended uniquely to a parallel class. For given two parallel lines, we can identify the point p on H where the meet; the remaining lines in the parallel class are those that also meet H at p.

Any collineation α of an affine geometry must map parallel lines to parallel lines, since it must map disjoint coplanar lines to disjoint coplanar lines. Thus α determines a bijection of the point set of H. It actually determines a collineation. To prove this we must find a way of recognising when the ‘points at infinity’ of three parallel classes are collinear. Suppose that we have three parallel classes. Choose a line line ℓ in the first. Since the parallel classes partition the points of the affine geometry, any point p on ℓ is also on a line from the second and the third parallel class. The points at infinity on these three lines are collinear, in H, if and only if the lines are coplanar. It follows that any collineation of an affine geometry determines a collineation of the hyperplane at infinity, and hence of the projective geometry. Because of the previous result, we can equally well view an affine geometry as a projective geometry $\mathcal{G}$ with a distinguished hyperplane. The points not on the hyperplane are the *affine points* and the lines of $\mathcal{G}$ not contained in H are the *affine lines*. It is important to realise that there are two different viewpoints available, and in the literature it is common to find an author shift from one to the other, without explicit warning.

There is a difficulty in providing a set of axioms for affine spaces, highlighted by the following. Consider the projective plane $PG(2,2)$. Removing a line from it gives the affine plane $PG(2,2)$ which has four points and six lines; each line has exactly two points on it. (Thus we could can identify its points and lines with the vertices and edges of the complete graph K_4 on four vertices.) This is a linear space but, unfortunately for us, it has rank four. Any set of three points is a subspace of rank three. More generally, any subset of the points of $AG(n,2)$ is a subspace of $AG(n,2)$ viewed as a linear space. However not all subspaces are flats. One set of axioms for affine spaces has been provided by H. Lenz. An incidence structure is an *affine space* if the following hold.

(a) Any two points lie on a unique line.

(b) Given any line ℓ and point p not on ℓ, there is a unique line ℓ' through p and disjoint from ℓ. (We say ℓ and ℓ' are parallel, and write $\ell \parallel \ell'$. Any line is parallel to itself.)

(c) If ℓ_0, ℓ_1 and ℓ_2 are lines such that $\ell_0 \parallel \ell_1$ and $\ell_1 \parallel \ell_2$ then $\ell_0 \parallel \ell_2$. (Or more clearly: parallelism is an equivalence relation on lines.)

(d) If $a \vee b$ and $c \vee d$ are parallel lines, and p is a point on $a \vee c$ distinct from a then $p \vee b$ intersects $c \vee d$.

(e) If a, b and c are three points, not all on one line, then there is a point d such that $a \vee b \parallel c \vee d$ and $a \vee c \parallel b \vee d$.

(f) Any line has at least two points.

It is not hard to show that all lines in an affine space must have the same number of points. This number is called the *order* of the space. If the order is at least three then the axiom (e) is implied by the other axioms. On the other hand, if all lines have two points then (d) is vacuously satisfied. Hence we are essentially treating separately the cases where the order is two, and where the order is at least three. Any line trivially satisfies the above set of axioms. If any two disjoint lines are parallel then we have an *affine plane*. These may be defined more simply as linear spaces which are not lines and have the property that, given any point p and line ℓ not on p, there is a unique line through p disjoint from ℓ. We can provide a simpler set of axioms for thick affine spaces. Call two lines in a linear space *strongly parallel* if they are disjoint and coplanar. Then the linear space $\mathcal{L}$ is an affine space if

(a) strong parallelism is an equivalence relation on the lines of $\mathcal{L}$,

(b) if p is a point, and ℓ is a line of $\mathcal{A}$, then there is a unique line through p strongly parallel to ℓ.

As with our first set of axioms, no mention is made of affine subspaces. However, in this case they are just the linear spaces. In the sequel, we will distinguish this set of axioms by referring to them as the "axioms for thick affine spaces". The first, official, set will be referred to as "Lenz's axioms".

2.6 Affine Spaces in Projective Space

We outline a proof that any thick affine space arises by obtained by deleting a hyperplane from a suitable projective plane.

2.6.1 Lemma. *Let $\mathcal{A}$ be a thick affine space with rank at least four. Let π be a plane in $\mathcal{A}$ and let D be a line intersecting, but not contained in π. Then the union of the point sets of those planes which contain D, and meet π in a line, is a subspace.*

Proof. Let W denote the union described. Since the subspace $D \vee \pi$ is the join of D and any line in π which does not meet π, no line in π which does not meet D can be coplanar with it. Hence no line in π is parallel to D. If x is point in W which is not on D then $x \vee D$ is a plane. Since $x \in W$, there is a plane containing x and D which meets π in a line. Thus $x \vee D$ must meet π in line, l say. As x is not on l, there is unique line, l' say, parallel to it through x. Since D is not parallel to l, it is not parallel to l'. Therefore D meets l'. We will denote the point of intersection of D with l' by $d(x)$. Now suppose that x and y are distinct points of W. We seek to show that any point on $x \vee y$ lies in W. There are unique lines through x and y parallel to D; since they lie in $x \vee D$ and $y \vee D$ respectively they meet π in points x' and y'. If u is a point on $x \vee y$ then the unique line through u parallel to D must intersect $x' \vee y'$. Hence u lies in the plane spanned by this point of intersection and D, and so $u \in W$. □

This lemma provides a very useful tool for working with affine spaces. We note some consequences.

2.6.2 Corollary. *Let π be a plane in the affine space $\mathcal{A}$ and let x and y be two points not on π such that $x \vee \pi = y \vee \pi$. If $x \vee y$ is disjoint from π then it is parallel to some line contained in π.*

Proof. Let p be a point in π. From the previous lemma we see that since $y \in x \vee \pi$, the plane spanned y and the line $x \vee p$ meets π in a line l. As l lies on π it is disjoint from $x \vee y$ and hence it is parallel to it. □

2.6.3 Corollary. *Let $\mathcal{A}$ be an affine space. If two planes in $\mathcal{A}$ have a point in common and are contained in subspace of rank four, they must have a line in common.*

Proof. Suppose that p is contained in the two planes σ and π. Let l be a line in σ which does not pass through p. As l is disjoint from π it is, by the previous corollary, parallel to a line l' in π. Let m be the line through p in σ parallel to l and let m' be the line in π parallel to p. Then

$$m' \parallel l',\ l' \parallel l,\ l \parallel m$$

and thus $m = m'$. Therefore $m \subseteq \sigma \cap \pi$. □

Let $\mathcal{A}$ be an affine geometry. We show how to embed it in a projective geometry. Assume that the rank of $\mathcal{A}$ is at least three. (If the rank is less than three, there is almost nothing to prove.) Let P be a set with cardinality equal to the number of parallel classes. We begin by adjoining P to the point set of $\mathcal{A}$. If a line of $\mathcal{A}$ lies in the i-th parallel class, we extend it by adding the i-th point of P. It is straightforward to show that each plane in $\mathcal{A}$ has now been extended to a projective plane. Each plane in $\mathcal{A}$ determines a set of parallel classes, and thus a subset of P. These subsets are defined to be lines of the extended geometry; the original lines will be referred to as *affine* lines if necessary. Two points a and b of $\mathcal{A}$ are collinear with a point p of P if and only the line $a \vee b$ is in the parallel class associated with p. With the additional points and lines as given, we now have a new incidence structure $\mathcal{P}$. We must verify that it is linear space. Let a and b be two points. If these both lie in $\mathcal{A}$ then there is a unique line through them. If $a \in \mathcal{A}$ and $b \in P$ then there is a unique line in the parallel class determined by b which passes through a. Finally, suppose that a and b are both in P. Let l be a line in the parallel class determined by a. If x is an affine point in l then there is unique line in the parallel class of b passing through it. With l, this line determines a plane which contains all the lines in b which meet l. This shows that each line l in a determines a unique plane. We claim that it is a projective space. This can be proved by showing that each plane in $\mathcal{P}$ is projective. The only difficult case is to verify that the planes contained in P are projective. Each plane of P corresponds to a subspace of $\mathcal{A}$ with rank four, so studying the planes of P is really studying these subspaces of $\mathcal{A}$. The planes contained in P are projective planes if every pair of lines in them intersect. Thus we must prove that if σ and π are two planes of $\mathcal{A}$ contained in a subspace of rank four, then there is line in σ parallel to π. There are two cases two consider. Suppose first that $\sigma \cap \pi = \emptyset$. Then, by Corollary 7.2, any line in σ is parallel to a line in π, and therefore there is a point in P lying on both the lines determined by σ and π. Suppose next that σ and π have a point in common. Then, by Corollary 7.3, these planes must have a line in common and so the parallel class containing it lies on the lines in P determined by them. This completes the proof that all affine spaces are projective spaces with a hyperplane removed. In the next chapter we will use our axiomatic characterisation of affine spaces to show that all projective spaces of rank at least four have the form $PG(n, \mathbb{F})$, that is, are projective spaces over some skew field. I do not know if the proof just given is any sense optimal, nor who introduced the axiom system we have used.

2.7 Characterising Affine Spaces by Planes

We have seen that a linear space with rank at least three is a projective geometry if and only if every plane in it is a projective plane. The corresponding result for affine geometries is more delicate and is due to Buekenhout.

2.7.1 Theorem. *Let $\mathcal{A}$ be a linear space with rank at least three. If each line has at least four points, and if all planes of $\mathcal{A}$ are affine planes, then $\mathcal{A}$ is an affine geometry.*

Proof. We verify that the axioms for thick affine spaces hold. Since the second of these axioms is a condition on planes, it is automatically satisfied. Thus we need only prove that parallelism is an equivalence relation on the lines of $\mathcal{A}$. If π is a plane and D a line meeting π in the point a, we define $W = W(\pi, D)$ to be the union of the point sets of the planes which contain D and meet π in a line.

Suppose $w \in W \setminus D$. The only plane containing w and D is $w \vee D$, hence the points of this plane must belong to W. In particular, it must meet π in a line l. Since w is not on l, there is a unique line m in $w \vee D$ through w and parallel to l. The line D meets l in a, and is therefore not parallel to it. Hence it is not parallel to m. Denote the point of intersection of m and D by $d(w)$. Note that if b is point on D, other than a or $d(w)$ then bw is a line in $w \vee D$ not parallel to l. Thus it must intersect l in a point.

Our next step is to show that W is a subspace. This means we must prove that if x and y are points in $W \setminus \pi$ then all points on xy lie in W. Suppose first that $xy \cap \pi = \emptyset$. Since the lines of $\mathcal{A}$ have at least four points on them, there is a point b on D distinct from a, $d(x)$ and $d(y)$. The line bx and by must meet π, in points x' and y' say. As xy and π are disjoint, $xy \cap x'y' = \emptyset$. Accordingly xy and $x'y'$ are parallel (they both lie in the plane $b \vee xy$). If u is point on xy then bu canot be parallel to $x'y'$ and so u is on a line joining b to a point of π. This implies that the plane $u \vee D$ meets π in two distinct points. Hence it is contained in W, and so $u \in W$, as required.

Assume next that xy meets π in a point, z say. Let σ be the plane $y \vee x \vee d(x)$. If $\sigma \cap \pi$ is a line then, since it is disjoint from $x \vee d(x)$, it is parallel to it. So, if u is a point distinct from x and y on xy then $u \vee d(x)$ cannot be parallel to $\sigma \cap \pi$. Accordingly $u \vee d(x)$ contains a point of π, implying as before that $u \vee D$ is in W. Hence $u \in W$. The only possibility remaining is that $\sigma \cap \pi$ is a point, in which case it is z. Assume u is a point distinct from x and y on xy. Since the line $z \vee d(x)$ has at least four points, and since there is only one line in σ parallel to $x \vee d(x)$ through u, there is a line through u meeting $x \vee d(x)$ and $z \vee d(x)$ in points x' and y' respectively. Now $x \vee d(x)$ is disjoint from π and therefore all points on it are in W. Also all points on $z \vee d(x)$ are in W. Hence x' and z' lie in W. Since z does not lie on $x'z'$, this line is disjoint from π. This shows that all points on it lie in W. We have finally shown that W is a subspace, and can now complete the proof of the theorem.

Suppose that l_1, l_2 and l_3 are lines in $\mathcal{A}$, with $l_1 \parallel l_2$ and $l_2 \parallel l_3$. Let π be the plane $l_1 \vee l_2$, let D be a line joining a point b on l_3 to a point a in l_2 and let $W = W(\pi, D)$. Since $b \in W$ and W is a subspace, the plane $b \vee l_1$ lies in W. In this plane there is a unique line through b parallel to l_1. Denote it by l_3'. As $l_3 \vee l_2$ meets π in l_2, we see that l_3 is disjoint from π. Similarly $l_3' \vee l_1$ meets π in l_1, and so l_3' is disjoint from π. The plane $a \vee l_3'$ is contained in W, and contains D. By the definition of W, any point of $a \vee l_3'$ lies in a plane which contains D and meets π in a line. This plane must be $a \vee l_3'$. Denote its line of intersection with π by l_2'. Since l_3' is disjoint from π, the lines l_2' and l_3' are parallel. If $l_2 = l_2'$ then l_3 and l_3' are two lines in $b \vee l_2$ intersecting in b and parallel to l_2. Hence they must be equal. If $l_2 \neq l_2'$ then l_2' must intersect l_1, in a point c say. But then l_1 and l_2' are lines in $c \vee l_3$ parallel to l_3. Therefore $l_1 = l_2'$, which is impossible since $a \in l_2'$ and $a \notin l_1$. Thus we are forced to conclude that $l_1 \parallel l_3$. □

The above proof is based in part on some notes of U. S. R. Murty. There are examples of linear spaces which are not affine geometries, but where every plane is affine. These were found by M. Hall; all lines in them have exactly three points.

Chapter 3

Collineations and Perspectivities

The main result of this chapter is a proof that all projective spaces of rank at least four, and all 'Desarguesian' planes, have the form $PG(n, \mathbb{F})$ for some field $\mathbb{F}$.

3.1 Collineations of Projective Spaces

A *collineation* of a linear space is a bijection ϕ of its point set such that $\phi(A)$ is a line if and only if A is. It is fairly easy to describe the collineations of the projective spaces over fields. Consider $PG(n, \mathbb{F})$, the points of which are the 1-dimensional subspaces of $V = V(n+1, \mathbb{F})$. Any invertible linear mapping of V maps 2-dimensional subspaces onto 2-dimensional subspaces, and hence induces a collineation of $PG(n, \mathbb{F})$. The set of all such collineations forms a group, called the projective linear group, and denoted by $PGL(n, \mathbb{F})$. There is however another class of collineations. Suppose τ is an automorphism of $\mathbb{F}$, e.g., if $\mathbb{F} = \mathbb{C}$ and τ maps a complex number to its complex conjugate. If $\alpha \in \mathbb{F}$, $x \in V$ and $\alpha^\tau \neq \alpha$ then

$$\alpha x^\tau = \alpha^\tau x^\tau \neq \alpha x^\tau .$$

Thus τ does not induce a linear mapping of V onto itself, but it does map subspaces to subspaces, and therefore does induce a collineation. If we apply any sequence of linear mappings and field automorphisms to $PG(n, \mathbb{F})$ then we can always obtain the same effect by applying a single linear mapping followed by a field automorphism (or a field automorphism then a linear mapping). The composition of a linear mapping and a field automorphism is called a *semi-linear* mapping. The set of all collineations obtained by composing linear mappings and field automorphisms is called the group of projective semi-linear transformations of $PG(n, \mathbb{F})$, and is denoted by $P\Gamma L(n+1, \mathbb{F})$. It contains $PGL(n, \mathbb{F})$ as a normal subgroup of index equal to $|\mathrm{Aut}(\mathbb{F})|$. (If $\mathbb{F}$ is finite of order p^m, where

p is prime, then $\mathrm{Aut}(\mathbb{F})$ is a cyclic group of order m generated by the mapping which sends an element x of $\mathbb{F}$ to x^p.) We can now state the "fundamental theorem of projective geometry".

3.1.1 Theorem. *Every collineation of $PG(n,\mathbb{F})$ lies in $P\Gamma L(n+1,\mathbb{F})$.*

Proof. Look it up, for example in Tsuzuku []. □

This theorem can be readily extended to cover collineations between distinct projective spaces over fields. These are all semi-linear too. It is even possible to describe all 'homomorphisms', that is, mappings from one projective space which take points to points and lines to lines, but which are not necessarily injective. (This requires the use of valuations of fields.) The most important property of $P\Gamma L(n+1,\mathbb{F})$ is that it is large. One way of making this more precise is as follows.

3.1.2 Theorem. . *The group $PGL(n,\mathbb{F})$ acts transitively on the set of all maximal flags of $PG(n-1,\mathbb{F})$.*

Proof. Exercise. □

Every invertible linear transformation of $V = V(n,\mathbb{F})$ determines a collineation of $PG(n-1,\mathbb{F})$. The group of all invertible linear transformations of V is denoted by $GL(n,\mathbb{F})$. This groups acts on $PG(n-1,\mathbb{F})$, but not faithfully—any linear transformation of the form cI, where $c \neq 0$, induces the identity collineation. (You will show as one of the exercises that all the linear transformations which induce the identity collineation are of this form.)

To compute the order of $PGL(n,\mathbb{F})$ when $\mathbb{F}$ is finite with order q, we first compute the order of $GL(n,\mathbb{F})$. This is just the number of non-singular $n \times n$ matrices over $\mathbb{F}$. We can construct such matrices one row at a time. The number of possible first rows is $q^n - 1$ and, in general, the number of possible $(k+1)$-th rows is the number of vectors not in the span of the first k rows, that is, it is $q^n - q^k$. Hence

$$|GL(n,q)| = \prod_{i=0}^{n-1}(q^n - q^i) = q^{\binom{n}{2}}(q-1)^n[n]!.$$

The number of maximal flags in $PG(n-1,\mathbb{F})$ is $[n]!$. Thus we deduce, using Theorem 1.2, that the subgroup G of $GL(n,\mathbb{F})$ fixing a flag must have order $q^{\binom{n}{2}}(q-1)^n$. This subgroup is isomorphic to the subgroup of all upper triangular matrices.

A *k-arc* in a projective geometry of rank n is a set of k points, no n of which lie in a hyperplane. To construct an $(n+1)$-arc in $PG(n-1,\mathbb{F})$, take a basis $x_1,\ldots,x_n$ of $V(n,\mathbb{F})$, together with a vector y of the form $\sum_i a_i x_i$, where none of the a_i are zero. The linear transformation which sends each vector x_i to $a_i x_i$ maps $\sum x_i$ to $\sum a_i x_i$. Hence the subgroup of $PGL(n,\mathbb{F})$ fixing each of $x_1,\ldots,x_n$ acts transitively on the set of points of the form $\sum a_i x_i$, where the

a_i are non-zero. It is also possible to show that a collineation of $PG(n-1,\mathbb{F})$ which fixes each point in an $(n+1)$-arc is the identity. (The proof of this is left as an exercise.) Together these statements imply that the subgroup of $GL(n,\mathbb{F})$ fixing each of $x_1,\dots,x_n$ acts regularly on the set of points of the form $\sum a_i x_i$, where the a_i are non-zero, and hence that it has order $(q-1)^n$. The subgroup of $P\Gamma L(n+1,\mathbb{F})$ fixing each point in an $(n+1)$-arc can be shown to be isomorphic to the automorphism group of the field $\mathbb{F}$. (See Hughes and Piper [].)

3.2 Perspectivities and Projections

A *perspectivity* of a projective geometry is a collineation which fixes each point in some fixed hyperplane (its *axis*), and each hyperplane through some point (its *centre*). The latter condition is equivalent to requiring that each line through some point be fixed, since every line is the intersection of the hyperplanes which contain it. While it it is clear that this is a reasonable definition, it is probably not clear why we would wish to consider collineations suffering these restrictions. However perspectivities arise very naturally. Let $\mathcal{G}$ be a projective geometry of rank four, and let H and K be two hyperplanes in it. Choose points p and q not contained in $H \cup K$. If $h \in H$, define $\phi_p(h)$ by

$$\phi_p(h) := (p \vee h) \cap K.$$

This works because H is a hyperplane, and so every line in $\mathcal{G}$ meets H. Similarly if $k \in K$ then we define $\psi_q(k)$ by

$$\psi_q(k) = (q \vee k) \cap H.$$

It is a routine exercise to show that ϕ_p is a collineation from H to K and ψ_q is a collineation from K to H. Hence their composition $\phi_p\psi_q$ is a collineation of H. (We made use of ϕ_p earlier in proving Theorem 4.3, that is, that all maximal subspaces of a projective geometry have the same rank.)

If $\mathcal{G}$ has rank n then the hyperplanes H and K meet in a subspace of rank $n-2$ and each point in this subspace is fixed by $\phi_p\psi_q$. All lines through the point $(p \vee q) \cap H$ are also left fixed by $\phi_p\psi_q$. As $H \cap K$ is a hyperplane in H, it follows that $\phi_p\psi_q$ is a perspectivity. We will make considerable use of these perspectivities in proving that all projective geometries of rank at least four arise as the 1- and 2-dimensional subspaces of a vector space. It is easy to provide a class of linear mappings of a vector space which induce perspectivities of the associated projective space $PG(n,\mathbb{F})$. They are known as *transvections*, and can be described as follows. Let $V = V(n,\mathbb{F})$ and let H be a hyperplane in V. A linear mapping τ of V is a transvection with *axis* H if $x\tau = x$ for all x in H, and $x\tau - x \in H$ for all x not in H. It is easy to construct transvections. Choose non-zero vectors h and a such that $(h,a) = 0$ and define $\tau_{h,a}$ by setting

$$x\tau_{h,a} = x - (x,h)a.$$

Then x is fixed by $\tau_{h,a}$ if and only if $(h,x) = 0$. Thus $\tau_{h,a}$ fixes all points of the hyperplane with equation $(h,x) = 0$. If x is not fixed by $\tau_{h,a}$ then $x\tau_{h,a} - x$

is a multiple of a and, since $(h, a) = 0$, it follows that a lies in the hyperplane of points fixed by $\tau_{h,a}$. As $\tau_{h,a}$ fixes a, it follows that it also fixes all the 2-dimensional subspaces $a \vee x$.

3.2.1 Lemma. *Let H be hyperplane in the projective geometry $\mathcal{G}$. If the collineation τ fixes all the points in H then it fixes all lines through some point of $\mathcal{G}$, and is therefore a perspectivity.*

Proof. Assume first that τ fixes some point c not in H and let l be a line through c. Then l must meet H in some point, x say. As $x \in H$, it is fixed by τ and thus τ fixes two distinct points of l. This implies that l is fixed by τ. Assume now that there are no points off H fixed by τ. Let p be a point not in H and let $l = p \vee p\tau$. Once again l must intersect H in some point, x say. As τ fixes x and maps p in l to $p\tau$ in l, it follows that it fixes l. Let q be a point not on H or l. The plane $\pi = q \vee l$ meets H in a line l' (why?). Since τ fixes the distinct lines l and l' from π, it also fixes π. This implies that $q\tau \in \pi$. Now $q\tau \neq q$, since $q \notin H$, and so the $q \vee q\tau$ is a line in π. Hence it intersects l' and, since $l' \subseteq H$, the point of intersection is fixed by τ. Therefore $q \vee q\tau$ is fixed by τ. The line $q \vee q\tau$ must intersect l in some point, c say. As $q \vee q\tau$ and l are both fixed by τ, so is c. Therefore $c \in H$ and so $c = H \cap l = x$. Thus we have shown that the lines $q \vee q\tau$, where $q \notin H$, all pass through the point c in H. From this it follows that all lines through c are fixed by τ. □

3.2.2 Corollary. *The set of perspectivities with axis H form a group.*

Proof. If τ is the product of two perspectivities with axis H, then it must fix all points in H. By the lemma, it is a perspectivity. □

Lemma 2.1 shows that perspectivities are the collineations which fix as many points as possible, and thus makes them more natural objects to study. By duality it implies that any collineation which fixes all hyperplanes on some point must fix all the points in some hyperplane. Note however that we cannot derive the lemma itself by appealing to duality, that is, by asserting that if τ fixes all points on some hyperplane then, by duality, it fixes all hyperplanes on some point. A perspectivity with its centre on its axis is often called an *elation*. If its centre is not on its axis it is a *homology*. (Classical geometry is full of strange terms.) From our remarks above, any transvection induces an elation. It can be shown that the perspectivities of $PG(n-1, \mathbb{F})$ all belong to $PG(n, \mathbb{F})$, and not just to $P\Gamma L(n, \mathbb{F})$. (In fact $PG(n, \mathbb{F})$ is generated by perspectivities in it.)

3.2.3 Corollary. *Let τ be a collineation fixing all points in the hyperplane H. If τ fixes no points off H it is an elation, if it fixes one point off H it is a homology and if it fixes two points off H it is the identity.*

Proof. Only the last claim needs proof. Suppose a and b are distinct points off H fixed by τ. If p is a third point, not in H, then τ fixes the point $H \cap pa$ as well as a. hence τ fixes pa and similarly it fixes pb. Therefore $p = pa \cap pb$ is fixed by τ. This shows that τ fixes all points not in H. □

3.3 Groups of Perspectivivities

In general the product of two perspectivities of a projective geometry need not be a perspectivity. There is an important exception to this.

3.3.1 Lemma. *Let τ_1 and τ_2 be perspectivities of the projective geometry $\mathcal{G}$ with common axis H. Then $\tau_1\tau_2$ is a perspectivity with axis H and centre on the line joining the centres of τ_1 and τ_2.*

Proof. Denote the respective centres of τ_1 and τ_2 by c_1 and c_2. Let c be the centre of $\tau_1\tau_2$ and let l be the line $c_1 \vee c_2$. We assume by way of contracdiction that c is not on l. As $c\tau_1\tau_2 = c$ it follows that $c\tau_1 = c\tau_2^{-1}$. Suppose that $c \neq c\tau_1$. Then $c\tau_1$ must lie on $c_1 \vee c$, since τ_1 fixes all lines through c_1. Similarly $c\tau_2^{-1}$ lies on $c_2 \vee c$. Hence

$$c_1 \vee c = c\tau_1 \vee c = c\tau_2^{-1} \vee c = c_2 \vee c,$$

implying that $c_2 \in c_1 \vee c$ and thus that $c \in l$. Thus c is fixed by both τ_1 and τ_2. If $c \notin H$ then we infer that c is the common centre of τ_1 and τ_2, whence we have $c = c_1 = c_2$. Thus we may assume that $c \in H$. Since l lies on c_1, it is fixed by τ_1 and, since it lies on c_2, it is also fixed by τ_2. Hence it is fixed by $\tau_1\tau_2$. If the centre of $\tau_1\tau_2$ is not on H then it must lie on l, as required. If $\tau_1\tau_2$ fixes no point off H then the proof of Lemma 2.1 shows that the centre of $\tau_1\tau_2$ is $l \cap H$. Thus we may assume that l lies in H. If our geometry has rank three then H must be equal to l, and so $c \in l$ as required. Thus we may assume that $\mathcal{G}$ has rank at least four, and that $c \notin l$. We show in this case that $\tau_1\tau_2$ is the identity. Let p be a point not in H. Then the plane $p \vee l$ is fixed by $\tau_1\tau_2$, and so is the line $p \vee c$ (because c is the centre of $\tau_1\tau_2$). As $c \notin l$, we see that p is the unique point of intersection of the line $p \vee c$ with the plane $p \vee l$. This shows that p must be fixed by $\tau_1\tau_2$. Since our choice of p off H was arbitrary, it follows that $\tau_1\tau_2$ is the identity collineation. □

One consequence of the previous lemma is that if Γ is a group of collineations of a projective geometry $\mathcal{G}$ then the perspectivities of $\mathcal{G}$ with axis H and with centres in the subspace F is a subgroup of Γ. In particular, the product of two elations with axis H is always a perspectivity with axis H and centre on H, that is, it is an elation. It is possible for the product of two homologies with axis H to be an elation—with its centre the point of intersection of H with the line through the centres of the homologies.

We now come to an important definition. Let p be a point and H a hyperplane in the projective geometry $\mathcal{G}$ and let Γ be a group of collineations of $\mathcal{G}$. Let $\Gamma(p, H)$ denote the subgroup of Γ formed by the perspectivities with centre p and axis H. We say that Γ is (p, H)-transitive if, for any line ℓ through p

which is not contained in H, the subgroup $\Gamma(p, H)$ acts transitively on the set of points of ℓ which are not on H. If $\mathrm{Aut}(\mathcal{G})$ is itself (p, H)-transitive then we say that $\mathcal{G}$ is (p, H)-transitive. This is a reasonable point to explain some group theoretic terms as well. If Γ is a permutation group acting on a set S and $x \in S$ then Γ_x is the subgroup of Γ formed by the permutations which fix x. Recall that the length of the orbit of x under the action of Γ is equal to the index of Γ_x in Γ. The group Γ is transitive if it has just one orbit on S. It acts *fixed-point freely* on S if the only element which fixes a point of S is the identity, that is, if Γ_x is the trivial subgroup for each element x in S. In this case each orbit of Γ on S will have length equal to $|\Gamma|$ (and so $|\Gamma|$ divides $|S|$ when everything is finite). Suppose that Γ is the group of all perspectivities of $\mathcal{G}$ with centre p and axis H. Let q be a point not in H and distinct from p. If an element γ of Γ fixes q then it is the identity. For since $q \notin H$ and since γ fixes each point in H it fixes all lines joining q to a point in H. But as H is a hyperplane, this means that it fixes all lines through q. Hence q must be the centre of γ, and so $q = p$. This contradiction shows that Γ must act fixed-point freely on the points of $\mathcal{G}$ not in $H \cup p$. In particular, for any line l, we see that Γ acts fixed-point freely on the points of $l \backslash p$ not in H. (Since $p \in l$, the line l must be fixed as a set by Γ.) Therefore if $\mathcal{G}$ is finite and $p \notin H$ then $|\Gamma|$ must divide $|l| - 2$, and if $p \in H$ then $|\Gamma|$ divides $|l| - 1$.

3.4 Desarguesian Projective Planes

Let $\mathcal{P}$ be a projective plane, with p a point and ℓ a line in it. The condition that $\mathcal{P}$ be (p, ℓ)-transitive can be expressed in a geometric form. A *triangle* in a projective plane is a set of three non-collinear points $\{a_1, a_2, a_3\}$, together with the lines $a \vee b$, $b \vee c$ and $c \vee a$. These lines are also known as the *sides* of the triangle. For convenience we will now begin to abbreviate expressions such as $a \vee b$ to ab. Two triangles $\{a_1, a_2, a_3\}$ and $\{b_1, b_2, b_3\}$ are said to be *in perspective from a point p* if the three lines a_1b_1, a_2b_2 and a_3b_3 all pass through p. They are in *perspective from a line* ℓ if the points $a_1a_2 \cap b_1b_2$, $a_2a_3 \cap b_2b_3$ and $a_3a_1 \cap b_3b_1$ all lie on ℓ. We have the following classical result, known as *Desargues' theorem.*

3.4.1 Theorem. *Let $\mathcal{P}$ be the projective plane $PG(2, \mathbb{F})$. If two triangles in $\mathcal{P}$ are in perspective from a point then they are in perspective from a line.*

Proof. Wait. □

A projective plane is (p, ℓ)-*Desarguesian* if, whenever two triangles $\{a_1, a_2, a_3\}$ and $\{b_1, b_2, b_3\}$ are in perspective from p and both $a_1a_2 \cap b_1b_2$ and $a_2a_3 \cap b_2b_3$ lie on ℓ, so does $a_3a_1 \cap b_3b_1$. We call a plane *Desarguesian* if it is (p, ℓ)-Desarguesian for all points p and lines ℓ. Since the projective planes over fields are all Desarguesian, by the previous theorem, this concept is quite natural. However we will see that a plane is Desarguesian if and only if it is of the form $PG(2, \mathbb{F})$ for some skew-field $\mathbb{F}$.

3.4.2 Theorem. *A projective plane is (p, ℓ)-transitive if and only if it is (p, ℓ)-Desarguesian.*

Proof. Suppose $\mathcal{P}$ is a (p, ℓ)-transitive plane. Let $\{a_1, a_2, a_3\}$ and $\{b_1, b_2, b_3\}$ be two triangles in perspective from p with both $a_1a_2 \cap b_1b_2$ and $a_2a_3 \cap b_2b_3$ lying on ℓ. By hypothesis, there is a perspectivity τ with centre p and axis ℓ which maps a_1 to b_1. Let x be the point $a_1a_2 \cap \ell$. Since $x\tau = x$, the perspectivity τ maps xa_1 onto xb_1. Now $xa_1 = a_1a_2$ and $xb_1 = b_1b_2$; thus τ maps a_1a_2 onto b_1b_2. Since the line pa_2 is fixed by τ, we deduce that $a_2 = pa_2 \cap a_1a_2$ is mapped onto $pa_2 \cap b_1b_2 = b_2$. A similar argument reveals that $a_3\tau = b_3$. Thus $(a_2a_3)\tau = b_2b_3$ and therefore $(a_2a_3 \cap \ell)\tau = b_2b_3 \cap \ell$. As τ fixes each point of ℓ, this implies that $a_2a_3 \cap \ell = b_2b_3 \cap \ell$ and hence that a_2a_3 and b_2b_3 meet at a point on ℓ. Thus our two triangles are (p, ℓ)-perspective.

We turn now to the slightly more difficult task of showing that if $\mathcal{P}$ is (p, ℓ)-Desarguesian then it is (p, ℓ)-transitive. Let x be a point distinct from p and not on ℓ and let y be a point of px distinct from p and not on ℓ. We need to construct a perspectivity with centre p and axis ℓ which sends x to y. If a is a point not on px or ℓ, define

$$a\tau := ((ax \cap \ell) \vee y) \cap pa$$

and if $a \in \ell$, set $a\tau$ equal to a. As thus defined, τ is a permutation of the point set of the affine plane obtained by deleting px from $\mathcal{P}$. We will prove that it is a collineation of this affine plane, and hence determines a collineation of $\mathcal{P}$ fixing px. Since $a\tau \in pa$, the mapping τ fixes the lines through p. Hence, if τ is a collineation then it is a perspectivity with centre and axis in the right place. Suppose that a and b are two distinct points of $\mathcal{P}$ not on px. If $b \in xa$ then $ax = ab$ and $((ax \cap \ell) \vee y) = ((ab \cap \ell) \vee y)$, implying that $b\tau$ is collinear with $y = x\tau$ and $a\tau$. Conversely, if $b\tau$ is collinear with y and $a\tau$ then b must be collinear with x and a. Thus we may assume that x, a and b are not collinear. Then $\{x, a, b\}$ and $\{y, a\tau, b\tau\}$ are two triangles in perspective from the point p. By construction xa meets $y \vee a\tau$ and xb meets $y \vee b\tau$ on ℓ. Therefore $a \vee b$ must meet $a\tau \vee b\tau$ on ℓ. Let u be a point on ab. Then, applying Desargues' theorem a second time, we deduce that au and $a\tau \vee u\tau$ meet on ℓ. Since $au = ab$, they must actually meet at $ab \cap \ell$. Therefore

$$a\tau \vee u\tau = a\tau \vee (\ell \cap ab) = a\tau \vee (\ell \cap (a\tau \vee b\tau)) = a\tau \vee b\tau$$

and so $u\tau$ is on $a\tau \vee b\tau$, as required. □

3.4.3 Lemma. *Let $\mathcal{G}$ be a projective geometry of rank at least four. Then all subspaces of rank three are Desarguesian projective planes.*

Proof. Let π be a plane in $\mathcal{G}$ and let p a point and ℓ a line in π. Let a and b be distinct points on a line in π through p, neither equal to p or on ℓ. Let σ be a second plane meeting π in ℓ and let v be a point not in $\pi \vee \sigma$ but not in π or σ. If $x \in \pi$ then $v \vee x$ must meet σ in a point. The mapping sending x

to $(v \vee x) \cap \sigma$ is collineation ϕ_v from π to σ. The line ba' is contained in the plane $p \vee a \vee v$, as is pv. Hence ba' meets pv in a point, w say. Note that w cannot lie in π or σ. Hence it determines a collineation ϕ_w from σ to π which maps a' to b. Both ϕ_v and ϕ_w fix each point in ℓ, and so their composition is a collineation of π which fixes each point of ℓ and maps a to b. This shows that π is a (p, ℓ)-transitive plane. As our choice of p and ℓ was arbitrary, it follows from the previous two results that all planes in $\mathcal{G}$ are Desarguesian.

3.4.4 Theorem. *A projective geometry with rank at least four is (p, H)-transitive for all points p and hyperplanes H.*

Proof. Let x and y be distinct points on a line through p, neither in H. If a is a point in $\mathcal{G}$ not on px define

$$a\tau = (((ax \cap H) \vee y) \cap pa.$$

(This is the same mapping we used in proving that a (p, ℓ)-transitive plane is (p, ℓ)-Desarguesian.) Let π be a plane through pa meeting H in a line. If $a \in \pi$ then $a\tau \in \pi$ and, from the proof of Theorem 4.2, it follows that τ induces a perspectivity on π with centre p and axis $\pi \cap H$. Thus if a and b are points not both on H and ab is coplanar with px, the image of ab under τ is a line. (The proof of Theorem 4.2 can also be used to show that tau can be extended to the points on px; we leave the details of this to the reader.) Suppose then that a and b are points not both on H and ab is not coplanar with px. The plane $x \vee ab$ meets H in a line ℓ, hence if $c \in ab$ then $c\tau$ lies in the intersection of the planes $p \vee ab$ and $y \vee \ell$. As $y \in px$, we have

$$y \vee \ell \subseteq px \vee \ell = px \vee ab.$$

Therefore $y \vee \ell$ is a hyperplane in $px \vee ab$ and so it meets $p \vee ab$ in a line. By construction, this line contains the image of ab under τ, and so we have shown that τ is a collineation. □

There are projective planes which are not Desarguesian, and so the restriction on the rank in the previous theorem cannot be removed. We will call an affine plane $\mathcal{P}^l$ Desarguesian if $\mathcal{P}$ is.

3.5 Translation Groups

Let H be a hyperplane in the projective geometry $\mathcal{G}$. (We assume that $\mathcal{G}$ has rank at least three.) The ordered pair $(\mathcal{G}, H)$ is an affine geometry and an elation of $\mathcal{G}$ with axis H and centre on H is called a *translation*. From Lemma 3.1, it follows that the set of all translations form a group. We are going to investigate the relation between the structure of $\mathcal{A}$ and this group. Some group theory must be introduced. A group Γ is *elementary abelian* if it is abelian and its non-identity elements all have the same order. If Γ is elementary abelian then so is any subgroup. As any element generates a cyclic group, and as the only elementary abelian cyclic groups are the groups of prime order, all non-zero elements of a finite elementary abelian group must have order p, for some prime p. The group itself thus has order p^n for some n. We will usually use multiplication to represent the group operation, and consequently refer to the 'identity element' rather than the 'zero element'. (There will be one important exception, when we consider endomorphisms.) If H and K are subsets of the group Γ then we define

$$HK = \{hk : h \in H, k \in K\}.$$

If H and K are subgroups and at least one of the two is normal then HK is a subgroup of Γ. If $S \subseteq \Gamma$ then $\langle S\rangle$ is the subgroup generated by S and $\langle 1\rangle$ is the trivial, or identity subgroup. Let $\mathcal{G}$ be a projective geometry and let H be a hyperplane in it. Let $\mathcal{A}$ be the affine geometry with H as the hyperplane at infinity. If F is a subspace of H then $T(F)$ is the group of all elations with axis H and centre in F. If we need to identify H explicitly we will write $T_H(F)$.

3.5.1 Lemma. *Let H be a hyperplane in the projective geometry $\mathcal{G}$. If p and q are distinct points on H such that $T(p)$ and $T(q)$ are both non-trivial then $T(H)$ is elementary abelian.*

Proof. Since a non-identity elation has a unique centre, $T(p) \cap T(q) = \langle 1\rangle$. Suppose that α and β are non-identity elements of $T(p)$ and $T(q)$ respectively. If l is a line through p then so is $l\beta^{-1}$. Hence the latter is fixed by α and

$$l\beta^{-1}\alpha\beta = l\beta^{-1}\beta = l.$$

This shows that $\beta^{-1}\alpha\beta \in T(p)$. In other words, $T(p)$ is normalised by the elements of $T(q)$. If $\beta^{-1}\alpha\beta \in T(p)$ then the commutator $\alpha^{-1}\beta^{-1}\alpha\beta$ must also lie in $T(p)$. A similar argument shows that $\alpha^{-1}\beta^{-1}\alpha \in T(q)$. Accordingly $\alpha^{-1}\beta^{-1}\alpha\beta$ also lies in $T(q)$. As $T(p) \cap T(q) = \langle 1\rangle$, it follows that $\alpha^{-1}\beta^{-1}\alpha = 1$. Consequently $\alpha\beta = \beta\alpha$. (In other words, two non-identity elations with the same axis and distinct centres commute.) We now show that $T(p)$ is abelian. Let α' be a second non-identity elementt of $T(p)$. Then $\alpha'\beta$ is an elation. If its centre is p then β must belong to $T(p)$. Thus its centre is not p. Arguing as before, but with $\alpha'\beta$ in place of β, we deduce that α and $\alpha'\beta$ commute. This implies in turn that α and α' commute. Finally, assume that α is an element of $T(p)$ with order m. If $\beta^p \neq 1$ then

$$(\alpha\beta)^p = \alpha^p\beta^p = \beta^p \in T(q). \tag{3.1}$$

Since $\alpha\beta$ is an elation with axis H, so is $(\alpha\beta)^p$, and (3.1) shows that its centre is q. Therefore the centre of $\alpha\beta$ is q and so $\alpha\beta \in T(q)$. Since $\beta \in T(q)$, we infer that α also lies in $T(q)$. This is impossible, and forces us to conclude that $\beta^p = 1$. Thus we have proved that two non-identity elations with distinct centres must have the same order. It is now trivial to show that $T(H)$ is elementary abelian. □

The group $T(p)$ may contain no elements of finite order, but in this case it is still elementary abelian.

3.5.2 Lemma. *Let H be a hyperplane in the projective geometry $\mathcal{G}$. If $\mathcal{G}$ is (p,H)-transitive and (q,H)-transitive then it is (r,H)-transitive for all points r on $p \vee q$.*

Proof. If $p = q$ there is nothing to prove, so assume they are not equal. Let r be a point on pq and let a and b be distinct points not on H and colinear with r. We construct an elation mapping a to b. The lines ab and pq are coplanar; let x be the point $pa \cap ab$. Since $\mathcal{G}$ is (p,H)-transitive, there is an element α of $T(p)$ which maps a to x. Similarly there is an element β of $T(q)$ mapping x to b. Hence the product $\alpha\beta$ maps a to b. It fixes r, and therefore it fixes the line $ra = ab$. Thus it is an elation with centre r. □

Any element of $T(p)T(q)$ is an elation with centre on $p \vee q$. Thus the proof of the lemma implies the following.

3.5.3 Corollary. *If $\mathcal{G}$ is (p,H)- and (q,H)-transitive then $T(p \vee q) = T(p)T(q)$.* □

3.6 Geometric Partitions

Assume now that $\mathcal{G}$ is a projective geometry which is (p,H)-transitive for all points p on the hyperplane H, e.g., any projective geometry with rank at least four, or any Desarguesian plane. Then $T(H)$ is an elementary abelian group and the subgroups $T(p)$, where $p \in H$, partition its non-identity elements. In fact $T(H)$, together with the subgroups $T(p)$, completely determines $\mathcal{G}$. The connection is quite simple: the elements of $T = T(H)$ correspond to the points of $\mathcal{G}\backslash H$ and the cosets of the subgroups $T(p)$ are the lines. The correspondence between points and elements of T arise as follows. Let o be a point not in H. We associate with the identity of T. If a is a second point not on H then there is a unique elation τ_a with axis H and centre $H \cap oa$ which maps o to a. Then the map $a \mapsto \tau_a$ is a bijection from T to the points of $\mathcal{G}^H$. If l is a line of $\mathcal{G}^H$ then then the affine points of l are an orbit of $T(l \cap H)$, and conversely, each such orbit is a line. This leads us naturally to conjecture that an elementary abelian group T, together with a collection of subgroups T_i $(i = 1, \dots, m)$ such that the sets $T_i\backslash 1$ partition $T\backslash 1$, determines an affine geometry. This conjecture is wrong, but easily fixed. Let T be an elementary abelian group. A collection of subgroups T_i $(i = 1, \dots, m)$ is a *geometric partition* of T if

(a) The sets $T_i \setminus 1$ partition $T \setminus 1$,

(b) $T_i \cap T_jT_k \neq \emptyset$ implies that $T_i \leqslant T_jT_k$.

A set of subgroups for which (a) holds is called a *partition* of T, although it is not quite. The partitions we have been studying are all geometric. For $T(p)T(q) = T(p \vee q)$ and so if $\tau \in T(r) \cap T(p)T(q)$ then r must lie on $p \vee q$

and so $T(r) \leqslant T(p)T(q)$. A geometric partition of an elementary abelian group determines an affine geometry $\mathcal{G}^H$. We take the affine points to be the elements of T and the lines to be the cosets of the subgroups T_i. This gives us a linear space. Showing that this is an affine geometry is left as an exercise.

3.6.1 Lemma. *Let T_i ($i = 1, \ldots, m$) be a geometric partition of the elementary abelian group T and let $\mathcal{A} = \mathcal{G}^H$ be the affine geometry it determines. If o is the point of $\mathcal{A}$ corresponding to the identity of T then any (o, H)-homology of $\mathcal{G}$ determines an automorphism of T which fixes each subgroup T_i, and conversely.*

Proof. Let α be an (o, H)-homology of $\mathcal{G}$. If $\tau \in T$, then we regard it as an elation of $\mathcal{G}$ and thus we can define $\tau^\alpha = \alpha^{-1}\tau\alpha$. Then τ^α fixes each point off H and the line joining o to the centre of τ. Hence, if $\tau \in T_i$, so is τ^α. As α is an element and T a subgroup of the collineation group of $\mathcal{A}$, the mapping $\tau \mapsto \tau^\alpha$ is an automorphism of T. The proof of the converse is a routine exercise. □

For the remainder of this section, we will represent the group operation in abelian groups by addition, rather than multiplication. This also means that the identity now becomes the zero element. If α and β are automorphisms of the abelian group T then we can define their sum $\alpha+\beta$ by setting $\tau^{\alpha+\beta}$ equal to $\tau^\alpha + \tau^\beta$, for all elements τ of T. This will not be an automorphism in general, but it is always an endomorphism of T. The endomorphisms of an abelian group form a ring with identity. We require one preliminary result.

3.6.2 Lemma. *Let T_i ($i = 1, \ldots, m$) be a geometric partition of the elementary abelian group T. If $T_k \leq T_i + T_j$ and $k \neq i$ then $T_i + T_j = T_i + T_k$.*

Proof. This can be proved geometrically, but we offer an alternative approach. We claim that

$$T_i + ((T_i + T_k) \cap T_j) = (T_i + T_k) \cap (T_i + T_j). \tag{3.2}$$

To prove this, note first that both terms on the left hand side are contained in the right hand side. Conversely, if u belongs to the right hand side then we can write it both as $x + y$ where $x \in T_i$ and $y \in T_j$, and as $x' + z$ where $x' \in T_i$ and $z \in T_k$. Since $x + y = x' + z$ we have $y = -x + x' + z \in (T_i + T_k)$ and so $y \in (T_i + T_k) \cap T$. If $T_k \leq T_i + T_j$ then the right hand side of (3.2) is equal to $T_i + T_k$ while, since the partition is geometric, the left hand side equals T_i or $T_i + T_j$. As $T_k \neq T_i$, this provs the lemma. □

3.6.3 Lemma. *Let T_i ($i = 1, \ldots, m$) be a geometric partition of the elementary abelian group T. Then the set of all endomorphisms of T which map each subgroup T_i into itself forms a skew field.*

Proof. Let K be the set of endomorphisms referred to. We show first that the elements of K are injective. Suppose that $\alpha \in K$ and $x\alpha = 0$ for some element

x of T. Assume that x is a non-zero element of T_1 and let y be a non-zero element of T_i for some i. Then

$$(x+y)\alpha = x\alpha + y\alpha = y\alpha$$

and therefore $(x+y)\alpha$ must lie in T_i, since $y\alpha$ does. On the other hand, $x+y$ cannot lie in T_i, and therefore $(x+y)\alpha = 0$. This shows that $y\alpha = 0$. As our choice of y in T_i was arbitrary, it follows that each element of T_i is mapped to zero and, as our choice of i was arbitrary, that $(T \setminus T_i)\alpha = 0$. Since $y\alpha = 0$, we may also reverse the role of x and y in the first step of our argument and hence deduce that $T_1\alpha = 0$. Thus we have proved that if α is not injective then it is the zero endomorphism.

We now show that the non-zero elements of K are surjective. Suppose that $v \in T_i + T_j$ and $\alpha \in K$. We prove that v is in the range of α. We may assume that $v \in T_i$. Choose a non-zero element u of T_j. Then $u\alpha \neq 0$ and we may also assume that $u\alpha - v \neq 0$. Then $u\alpha - v$ must lie in some subgroup T_k and T_k must be contained in $T_i + T_j$. Since $T_k + T_j = T_i + T_j$, we see that T_k is a complete set of coset representatives for T_j in $T_i + T_j$ and so $T_k + u$ must contain a non-zero element w of T_i. Now $w - u \in T_k$ and therefore $(w-u)\alpha \in T_k$. As $u\alpha - v \in T_k$ we see that $w\alpha - v \in T_k$. On the other hand, v and w belong to T_i and so $w\alpha - v \in T_i$. Hence $w\alpha - v \in T_i \cap T_k = 0$. Consequently v lies in the range of α. We have now proved that any non-zero element of K is bijective. It follows that all non-zero elements of K are invertible, and hence that it is a skew field. □

A famous result due to Wedderburn asserts that all finite skew fields are fields. It is useful to keep this in mind. It is a fairly trivial exercise to show that any endomorphism of a geometric partition induces a homology of the corresponding projective geometry.

3.7 The Climax

The following result will enable us to characterise all projective geometries of rank at least four, and all Desarguesian projective planes.

3.7.1 Theorem. *Let $\mathcal{G}$ be a projective geometry of rank at least two, and let H be a hyperplane such that $\mathcal{G}$ is (p,H)-transitive for all points p in H. Then if $\mathcal{G}$ is (o,H)-transitive for some point o not in H, it is isomorphic to $PG(n,\mathbb{F})$ for some skew field $\mathbb{F}$.*

Proof. Let $T = T(H)$ and let K be the skew field of endomorphisms of the geometric partition determined by the subgroups $T(p)$, where $p \in H$. The non-zero elements of K form a group isomorphic to the group of all homologies of $\mathcal{G}$ with axis H and centre some point o off H. Since K is a skew field, we can view T as a vector space (over K) and the subgroups $T(p)$ as subspaces. As T acts transitively on the points of $\mathcal{G}$ not in H, it follows that $\mathcal{G}$ is (o,H)-transitive. This implies that $K \setminus 0$ acts transitively on the non-identity elements of $T(p)$, and hence that $T(p)$ is 1-dimensional subspace of T. Consequently the affine geometry $\mathcal{G}^H$ has as its points the elements of the vector space T, and as lines the cosets of the 1-dimensional subspaces of T. Hence it is $AG(n, K)$, for some n. This completes the proof. □

We showed earlier that every projective geometry $\mathcal{G}$ with rank at least four was (p, H)-transitive for any hyperplane H and any point p. Hence we obtain:

3.7.2 Corollary. *A projective geometry of rank at least four has the form $PG(n, \mathbb{F})$ for some skew field $\mathbb{F}$.* □

Similarly we have the following.

3.7.3 Corollary. *A Desarguesian projective plane has the form $PG(2, \mathbb{F})$ for some skew field $\mathbb{F}$.* □

If $\mathcal{P}$ is a projective plane which is (p, l)-transitive for all points on some line l then the affine plane $\mathcal{P}^l$ is called a *translation plane*. Translation planes which are not Desarguesian do exist, and some will be found in the next chapter.

Chapter 4

Spreads and Planes

We are going to construct some non-Desarguesian translation planes. This will make extensive use of the theory developed in the previous chapter.

4.1 Spreads

Every projective geometry which is (p, H)-transitive for all points p on some hyperplane H gives rise, as we have seen, to a geometric partition of an abelian group T. The ring of endomorphisms of this partition is a skew field K. Hence T is a vector space over K and the subgroups $T(p)$ are subspaces. These all have the same dimension over K. To see this note that $T(p)T(q)$ contains elements not in $T(p) \cup T(q)$ and so there is a point r, not equal to p or q, such that $T(r) \subseteq T(p)T(q)$. Since $T(p)T(r) = T(q)T(r)$ and $T(p)$, $T(q)$ and $T(r)$ are disjoint, it follows that $T(p)$ and $T(q)$ must have the same dimension. Our claim follows easily from this. It is not hard to see that the original geometry is a plane if and only if $T = T(p)T(q)$ for any pair of distinct points p and q. A geometric partition with this property is called a *spread*.

Since projective geometries with rank at least four are all of the form $PG(n, \mathbb{F})$, we no longer have much reason to bother working with geometric partitions in general. However spreads remain objects of considerable interest. Spreads can be defined conveniently as follows. Let $V = V(2n, \mathbb{F})$ be a vector space over the skew field $\mathbb{F}$. A spread is set of n-dimensional subspaces of V which partitions the non-zero elements of V. These subspaces are often referred as the *components* of the spread. Every spread determines a translation plane, on which the vector space V acts as a group of translations. The ring consisting of the endomorphisms of V which fix each component is the *kernel* of the spread (or of the plane it determines). As we have seen, it is a skew field, which necessarily contains $\mathbb{F}$ in its centre. The points of the affine plane can be identified with the elements of V and the lines are then the cosets (in V) of the components of $\mathcal{S}$. The point corresponding to the zero of V will be denoted by o.

4.1.1 Lemma. *Let $\mathcal{A}$ be the affine plane determined by the spread $\mathcal{S}$ of $V(2n, \mathbb{F})$ and let K be its kernel. Then the collineations of $\mathcal{A}$ which fix o are induced by the semilinear mappings of V which map the components of $\mathcal{S}$ onto themselves.*

Proof. Let α be a collineation of $\mathcal{A}$ fixing o. If $v \in V$, define the mapping τ_u on the points of $\mathcal{A}$ by

$$\tau_u(x) = x + u.$$

A routine check shows that this is a translation of $\mathcal{A}$. It is also easy to show that if τ is a translation then so is $\alpha^{-1}\tau\alpha$. Hence the mapping

$$\tau \mapsto \alpha^{-1}\tau\alpha$$

is an automorphism of the group of translations of $\mathcal{A}$. Thus it induces an additive mapping of V. Similarly we see that if β is a homology of $\mathcal{A}$ with centre o and axis the line at infinity then so is $\alpha^{-1}\beta\alpha$. The group formed by these homologies is isomorphic to the multiplicative group formed by the non-zero elements of K. AS V is a vector space over K, it follows that α induces a semilinear mapping of V. That is, it can be represented as the composition of a linear mapping and an automorphism of the skew field K. The converse is straightforward. □

Lemma 1.1 can be extended without thought to isomorphisms between translation planes. The next result is an important tool for working with spreads.

4.1.2 Lemma. *Let $V = V(2n, \mathbb{F})$ and let X_1, X_2, X_3 and Y_1, Y_2 and Y_3 be subspaces such that*

$$X_1 \oplus X_2 = X_2 \oplus X_3 = X_3 \oplus X_1 = V$$

and

$$Y_1 \oplus Y_2 = Y_2 \oplus Y_3 = Y_3 \oplus Y_1 = V.$$

Then there is linear mapping σ in $GL(V)$ such $X_i\sigma = Y_i$.

Proof. Our hypothesis implies all six subspaces have dimension n and that X_1 and X_2 are disjoint (well, excepting zero). Each subspace can be represented as the row space of an $n \times 2n$ matrix over $\mathbb{F}$. There is an element α of $GL(V)$ sending X_1 to the subspace equal to the row space of the $n \times 2n$ matrix $[I\,0]$ and X_2 to the row space of $[0\,I]$. Suppose that $X_3\alpha$ is the row space of $[A\,B]$. Since X_3 is disjoint from X_1, the matrix

$$\begin{pmatrix} I & 0 \\ A & B \end{pmatrix}$$

must be non-singular. This implies that B must be non-singular. As X_3 is disjoint from X_2, we deduce similarly that A is non-singular. Let β be the element

$$\begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix}$$

of $GL(V)$. Then $[A\,B]\beta = [I\,I]$, while the row spaces of $[I\,0]$ and $[0\,I]$ are both fixed by β. (Do not forget that $[I0]$ and $[A^{-1}\,0]$ have the same row space.) Thus there is an element of $GL(V)$ which sends X_1, X_2 and X_3 respectively to the row spaces of the matrices $[I\,0]$, $[0\,I]$ and $[I\,I]$. The lemma follows at once from this. □

The representation of the components of a spread in $V(2n, \mathbb{F})$ by the row spaces of $n \times 2n$ matrices is very useful. By virtue of the previous lemma, we may assume that a given spread contains the rows spaces of the matrices $[0\,I]$ and $[I\,0]$. Thus any third subspace is the row space of a matrix $[A\,B]$ where A and B are non-singular. As the row space of $[A\,B]$ and $[I\,A^{-1}B]$ are equal, this means each of the remaining subspaces can be specified by a $n \times n$ invertible matrix. (In this case, $A^{-1}B$.) The condition that the row spaces of $[I\,A]$ and $[I\,B]$ be disjoint is equivalent to the condition that the matrix

$$\begin{pmatrix} I & A \\ I & B \end{pmatrix}$$

be non-singular. This is equivalent to requiring that $B - A$ be non-singular, since this is the determinant of the above matrix. If A and B are elements of $GL(U)$ then $B - A$ is invertible if and only if $(I - B^{-1}A)$ is, and the latter holds if and only if there is no non-zero vector u such that $B^{-1}Au = u$. Thus $B - A$ is invertible if and only if $I - B^{-1}A$ acts fixed point freely on the non-zero elements of U. We will find that it is sometimes more convenient to verify that $A^{-1}B$ acts fixed-point freely than to show that $A - B$ is invertible. If σ is an $n \times n$ matrix then the row space of the matrix $[I\,\sigma]$ will be denoted by $V(\sigma)$. The row space of $[0\,I]$ will be denoted by $V(\infty)$.

4.1.3 Theorem. *Let $V = U \oplus U$ be a $2n$-dimensional vector space over $\mathbb{F}$. Then a spread of V is equivalent to a set Σ of elements of $GL(U)$, indexed by the non-zero elements of U, such that the difference of any two elements of Σ is invertible.*

Proof. Suppose we are given the set Σ. Then the subspaces $V(\infty)$, $V(0)$ and $V(\sigma)$ where $\sigma \in \Sigma$ are pairwise skew. To show that they form a spread we must verify that if (u, v) is a non-zero vector in V then it lies in one of these subspaces. If $u = 0$ or $v = 0$ then this is immediate. Consider the vectors $(u, u\sigma)$, where σ ranges over the elements of Σ. Since these act fixed-point freely on U, the vectors we obtain are all distinct. We obtain all vectors with first 'coordinate' u if and only if $|\Sigma| = |U \setminus 0|$. □

Theorem 1.3 provides us with a compact representation of a spread. Note that different spreads can give rise to the same translation plane. If α is the matrix

$$\begin{pmatrix} W & X \\ Y & Z \end{pmatrix}$$

in $GL(U \oplus U)$ then α maps the row space of $[I\,A]$ to the row space of $[W + AY\;X + AZ]$. If $W + AY$ is invertible this shows that the subspace parameterised by A is mapped to the subspace parameterised by $(W+AY)^{-1}(X+AZ)$. If α fixes $[I\,0]$ and $[0\,I]$ then both X and Y must zero. If $\Sigma\alpha = \Sigma$ then α induces a collineation of the affine plane determined by Σ.

4.2 Collineations of Translation Planes

Let $V = U \oplus U$ and let Σ be a subset of $GL(U)$ determining a spread $\mathcal{S}$ of V. Let $\mathcal{A}$ be the affine plane belonging to $\mathcal{S}$. The subspaces $V(\sigma)$ are the lines through the point $o = (0,0)$ in $\mathcal{A}$. The elements of V can all be written in the form (u,v), where u and v belong to U. Then

$$V(\infty) = \{(0,u) : u \in U\}$$

and

$$V(\sigma) = \{(u, u\sigma) : u \in U\}$$

for any element σ in $\Sigma \cup 0$. Since U is a vector space over the kernel K of $\mathcal{S}$, it follows that any non-identity automorphism of K must act non-trivially on it. (That is, it cannot fix each element of U.) From this it follows in turn that any perspectivity of $\mathcal{A}$ fixing o must be induced by a linear mapping of V, and not just a semilinear one. We consider the line at infinity l_∞ in $\mathcal{A}$ as a distinguished line, rather than as a missing line. Let (0) and (∞) be the points at which $V(0)$ and $V(\infty)$ respectively meet l_∞.

4.2.1 Theorem. *The set $\{\sigma \in GL(U) : \sigma\Sigma = \Sigma\}$ is a group, and is isomorphic to the group of homologies of $\mathcal{A}$ with centre (0) and axis $V(\infty)$. The set $\{\sigma \in GL(U) : \Sigma\sigma = \Sigma\}$ is a also a group, and is isomorphic to the group of homologies of $\mathcal{A}$ with centre (∞) and axis $V(0)$.*

Proof. Suppose that δ' is a homology of $\mathcal{A}$ with centre (0) and axis $V(\infty)$. Since δ' fixes each point on the line $V(\infty)$, it is induced by a linear mapping. Since δ' fixes the lines $V(0)$ and $V(\infty)$, it must map (u,v) to $(u\delta, v\gamma)$ for some elements δ and γ of $GL(U)$. (This follows from one of the remarks at the end of the previous section.) Since δ' fixes each point on $V(\infty)$, we must have $\gamma = 1$. Suppose that δ' maps $V(\sigma)$ to $V(\tau)$. Then

$$(u, u\sigma)\delta' = (u\delta, u\sigma)$$

and therefore $\delta^{-1}\sigma = \tau$. From this we infer that $\delta^{-1}\Sigma = \Sigma$, and so the first part of the lemma is proved. The second part follows similarly. The converse is routine. □

4.2.2 Corollary. *If Σ contains the identity of $GL(U)$ and the plane it determines is Desarguesian, then Σ is a group.*

Proof. If $\mathcal{A} = \mathcal{P}^l$ is Desarguesian then it is (p, H)-transitive for all points p and lines H. By the previous lemma, it follows that

$$\{\sigma \in GL(U) : \sigma\Sigma = \Sigma\}$$

has the same cardinality as Σ. Since $I \in \Sigma$, we see that if $\sigma\Sigma = \Sigma$ then σ must belong to Σ. Consequently Σ is closed under multiplication. As it consists of invertible matrices and contains the identity matrix, it is therefore a group. □

The group of homologies with centre (0) and axis $V(\infty)$ in the previous lemma has the same cardinality as Σ. This implies our claim immediately.

The converse to this corollary is false. (See the next section.) There is an analog of Lemma 2.1 for elations.

4.2.3 Lemma. *Let $\Sigma_0 = \Sigma \cup 0$. The set*

$$\{\sigma \in \Sigma_0 : \sigma + \Sigma_0 = \Sigma_0\}$$

is an abelian group, and is isomorphic to the group of elations with centre (∞) and axis $V(\infty)$.

Proof. If α is represented by the matrix

$$\begin{pmatrix} W & X \\ Y & Z \end{pmatrix}$$

and $(u, v) \in V$ then

$$(u, v)\alpha = (uW + vY, uX + vZ).$$

If $(0, v) \in V(\infty)$ then $(0, v)\alpha = (Yv, Zv)$. Hence if each point on $V(\infty)$ is fixed by α then $Y = 0$ and $Z = I$. If α also fixes all lines through (∞) then it must fix the cosets of $V(\infty)$. The elements of a typical coset of $V(\infty)$ have the form $(a, b + v)$, where v ranges over the elements of U. Now

$$(a, b + v)\alpha = (aW, aX + b + v)$$

and so if α fixes the lines parallel to $V(\infty)$ then $W = I$. Consequently, if α is an elation with centre (∞) and axis $V(\infty)$ and $\sigma \in \Sigma$ then

$$(u, u\sigma)\alpha = (u, uX + u\sigma)$$

As α is a collineation fixing o, it maps $V(\sigma)$ to $V(\tau)$ for some τ in Σ, or to $V(0)$. Therefore $X + \Sigma_0 = \Sigma_0$. Thus we have shown that the elations with centre (∞) and axis $V(\infty)$ correspond to elements $\sigma \in \Sigma$ such that $\sigma + \Sigma_0 = \Sigma_0$. The proof of the converse is routine. □

4.2.4 Corollary. *If the plane $\mathcal{P}$ determined by Σ is Desarguesian then Σ_0 is a skew field.*

Proof. Since $\mathcal{P}$ is (p,l)-transitive for all points p and lines l, we deduce from Lemma 2.1 that Σ is a group and from Lemma 2.3 that Σ_0 is a group under addition. If $\sigma \in \Sigma$ then $\sigma^{-1}\Sigma = \Sigma$, implying that $I = \sigma^{-1}\sigma \in \Sigma$. As both addition and multiplication are the standard matrix operations, the usual associative and distributive laws hold. Therefore Σ_0 is a skew field. □

4.2.5 Lemma. *If, for all elements σ and τ of Σ we have $\sigma\tau = \tau\sigma$ then Σ_0 is a field and the plane determined by Σ is Desarguesian.*

Proof. Suppose that α is an element of $GL(U)$ which commutes with each element of Σ. Then the map sending $(u, u\sigma)$ to

$$(u\alpha, u\sigma\alpha) = (u\alpha, u\alpha\sigma)$$

fixes each component of the spread $\mathcal{S}$ and hence it must lie in its kernel. Denote this by K. The hypothesis of the lemma thus implies that Σ is a commutative subset of $K \setminus 0$. The elements of Σ determine distinct homologies of the plane determined by the spread, with centre o and axis l_∞. Hence the plane must be Desarguesian (by Theorem 2.7.1) and Σ must coincide with $K \setminus 0$. □

4.3 Some Non-Desarguesian Planes

We propose to construct non-Desarguesian planes of order 9 and 16. Let U be a vector space over $\mathbb{F}$ and let Σ be a subset of $GL(U)$ determining a spread $\mathcal{S}$ of $V = U \oplus U$. As customary, we assume that $V(0)$ and $V(\infty)$ are components of $\mathcal{S}$. The plane determined by $\mathcal{S}$ is a *nearfield plane* if Σ is a group. (Thus Desarguesian planes are nearfield planes.)

First we construct a plane of order nine. Consider the group $SL(2,3)$ of 2×2 matrices over $GF(3)$ with determinant 1. Let U be the 2-dimensional vector space over $GF(3)$. We take Σ to be a Sylow 2-subgroup of $SL(2,3)$. Since $SL(2,3)$ has order 24, this means Σ has the right size. There is also no question that its elements are invertible.

To show that Σ determines a spread, we first show that 2-elements of $SL(2,3)$ act fixed-point freely on U. Suppose that $\alpha^2 = 1$. If $\alpha = \binom{ab}{cd}$ then the off-diagonal entries of σ^2 are $b(a+d)$ and $c(a+d)$. Hence either $b = c = 0$ or $a + d = 0$. In the first case, since $\det\alpha = 1$, we deduce that $\alpha = \pm I$. Otherwise it follows that α has the form

$$\begin{pmatrix} a & b \\ -(1+a^2)/b & -a \end{pmatrix}$$

whence a simple calculation shows that $\alpha^2 = -1$. Thus -1 is the only involution in $SL(2,3)$. As it acts fixed-point freely on U, all 2-elements of $SL(2,3)$ must act fixed-point freely. If σ and τ belong to Σ then $\sigma^{-1}\tau$ is a 2-element, and so acts fixed point freely on U. Hence $\sigma - \tau$ is invertible and therefore Σ determines a spread of $U \oplus U$. Since Σ is not commutative, the plane we obtain is not Desarguesian.

Our second plane needs more work. Consider the projective plane over $GF(2)$. If we number its points 1 through 7, its lines may be taken to be

$$123,\ 145,\ 167,\ 246,\ 257,\ 347,\ 356.$$

Each line gives us two 3-cycles belonging to the alternating group A_7. (For example the line 257 produces (257) and (275).) Let Σ be the set formed by these fourteen 3-cycles, together with the identity. Let X be the 4-dimensional vector space over $GF(2)$. We claim that A_7 can be viewed as a subgroup of $GL(4,2)$ acting transitively on the 15 non-zero elements of X. The proof of this is given in the next section. We prove that if σ and τ are elements of Σ then $\sigma^{-1}\tau$ acts fixed-point freely on the non-zero vectors of X. A routine check shows that $\sigma^{-1}\tau$ is either a 3-cycle or a 5-cycle. If x is a non-zero vector in X then the subgroup of A_8 leaving it fixed has order $8!/30 = 21 \cdot 2^6$. Thus this subsgroup contains no elements of order 5, and so all elements of order 5 in A_8 must act fixed-point freely. Suppose then that $\theta = \sigma^{-1}\tau$ is 3-cycle in A_8. Then there is a 5-cycle ϕ which commutes with θ. If x is non-zero vector fixed by θ then

$$x\phi\theta = x\theta\phi = x\phi$$

and so $x\phi$ is also fixed by θ. This shows that the number of non-zero vectors fixed by θ is divisible by 5. As θ has order three, the number of non-zero vectors not fixed by it is divisible by three. This implies that θ cannot fix 5 or 10 vectors, and hence that it must have 15 fixed points, that is, it is the identity element. Thus we have now shown that Σ determines a spread of $X \oplus X$. The resulting plane is not a nearfield plane, for then Σ would be a group of order 15. The only group of order 15 is cyclic, and hence abelian. But the Sylow 2-subgroups of $SL(2,3)$ are isomorphic to the quaternion group, which is not abelian. The plane we have constructed is called the *Lorimer-Rahilly* plane. Note that the collineation group of the plane over $GF(2)$ induces a group of collineations of the new plane fixing $V(0)$, $V(1)$ and $V(\infty)$, and acting transitively on the remaining components.

4.4 Alt(8) and $GL(4,2)$ are Isomorphic

We outline a proof that A_8 is isomorphic to $GL(4,2)$. Let S be the set $\{0,1,\ldots,7\}$. There are 35 partitions of S into two sets of size four and since S_8 acts on S, it also acts on this set of partitions. Any partition can be described by giving the elements of the component containing 1. Let Ω be the set of all 35 triples from $S\backslash 0$. It is not hard to check that A_7 acts transitively on Ω. A set of seven triples from Ω will be called a *heptad* if it has the property that every pair of triples from it intersect in precisely one point, and there is no point in all seven. We say that a set of triples are *concurrent* if there is some point common to them all, and the intersection of any two of them is this common point. A *star* is a set of three concurrent triples. The remainder of the argument is broken up into a number of separate claims.

4.4.1 Claim. *No two distinct heptads have three non-concurrent triples in common.*

It is only necessary to check that for one set of three non-concurrent triples, there is a unique heptad containing them.

4.4.2 Claim. *Each star is contained in exactly two heptads.*

Without loss of generality we may take our star to be 123, 145 and 167. By a routine calculation one finds that there are two heptads containing this star:

123	123
145	145
167	167
246	247
257	256
347	346
356	357

Note that the second of these heptads can be obtained from the first by applying the permutation (67) to each of its triples.

4.4.3 Claim. *There are exactly 30 heptads.*

There are 15 stars on each point, thus we obtain 210 pairs consisting of a star and a heptad containing it. As each heptad contains exactly 7 stars, it follows that there must be 30 heptads.

4.4.4 Claim. *Any two heptads have 0, 1 or 3 triples in common.*

If two heptads have four (or more) triples in common then they have three non-concurrent triples in common. Hence two heptads can have at most three triples in common. If two triples meet in precisely one point, there is a unique third triple concurrent with them. Any heptad containing the first two triples must contain the third. (Why?)

4.4.5 Claim. *The automorphism group of a heptad has order 168, and consists of even permutations.*

First we note that Sym(7) acts transitively on the set of heptads. As there are 30 heptads, we deduce that the subgroup of Sym(7) fixing a heptad must have order 168. Now consider the first of our heptads above. It is mapped onto itself by the permutations (24)(35), (2435)(67), (246)(357) and (1243675). The first two of these generate a group of order 8. Hence the group generated by these four permutations has order divisible by 8, 3 and 7. Since its order must divide 168, we deduce that the given permutations in fact generate the full automorphism group of the heptad.

4.4.6 Claim. *The heptads form two orbits of length 15 under the action of A_7. Any two heptads in the same orbit have exactly one triple in common.*

Since the subgroup of A_7 fixing a heptad has order 168, the number of heptads in an orbit is 15. Let Π denote the first of the heptads above. The permutations (123), (132) and (145) lie in A_7 and map Π onto three distinct heptads, having exactly one triple in common with Π. (Check it!) From each triple in Π we obtain two 3-cycles in A_7, hence we infer that there are 14 heptads in the same orbit as Π under A_7 and with exactly one triple in common with Π. Since there are only 15 heptads in an A_7 orbit, and since all heptads in an A_7 orbit are equivalent, it follows that any two heptads in such an orbit have exactly one triple in common.

4.4.7 Claim. *Each triple from Ω lies in exactly six heptads, three from each A_7 orbit.*

Simple counting.

4.4.8 Claim. *A heptad in one A_7 orbit meets seven heptads from the other in a star, and is disjoint from the remaining eight.*

More counting.

Now we construct a linear space. Choose one orbit of heptads under the action of A_7, and call its elements points. Let the triples be the lines, and say that a point is on a line if the correponding heptad contains the triple. The elements of the second orbit of heptads under A_7 determine subspaces of rank three, each isomorphic to a projective plane. It is now an exercise to show that there are no other non-trivial subspaces, and thus we have a linear space of rank four, with all subspaces of rank three being projective planes. Hence our linear space is a projective geometry, of rank four. Since its lines all have cardinality three, it must be the projective space of rank four over $GF(2)$. As $GF(2)$ has no automorphisms, the collineation group of our linear space consists entirely of linear mappings; hence it is isomorphic to $GL(4,2)$. (Note that we have just used the characterisation of projective geometries as linear spaces with all subspaces of rank three being projective planes, the fact that projective geometries of rank at least four are all of the form $PG(n,\mathbb{F})$ and the fundamental theorem of projective geometry, i.e., that the collineations of $PG(n,\mathbb{F})$ are semilinear mappings.) Our argument has thus revealed that A_7 is isomorphic to a subgroup of $GL(4,2)$. A direct computation reveals that it has index eight.

With a little bit of group theory it now possible to show that $GL(4,2)$ is isomorphic to A_8. We outline an alternative approach. Let Φ be the set of all partitions of S into two sets of size four. These sets can be described by giving the three elements of $S\backslash 0$ which lie in the same component of the partition as 0. Since S_8 acts on S, we thus obtain an action of S_8 on the 35 triples in Ω. This action does not preserve the cardinality of the intersection of triples. However if two triples meet in exactly one point then so do their images. (Because two triples meet in one point if and only if the meet of the corresponding partitions is a partition of S into four pairs.) Hence the action of S_8 on Ω does preserve heptads. More work shows that, in this action, A_8 and A_7 have the same orbits on heptads. Thus A_8 is isomorphic to a subgroup of $GL(4,2)$, and hence to $GL(4,2)$.

4.5 Moufang Planes

A line l in a projective plane $\mathcal{P}$ is a *translation line* if $\mathcal{P}$ is (p,l)-transitive for all points p on l, that is, if $\mathcal{P}^l$ is a translation plane. We call p a *translation point* if $\mathcal{P}$ is (p,l)-transitive for all lines l on it. From Lemma 2.3.1, we know that if $\mathcal{P}$ is (p,l)-transitive and (q,l)-transitive for distinct points p and q on l then l is translation line. Dually, if $\mathcal{P}$ is (p,l)- and (p,m)-transitive for two lines l and m through p then p is a translation point. The existence of more than one translation line (or point) in a projective plane is a strong restriction on its structure. The first conseequence is the following.

4.5.1 Lemma. *If l and m are translation lines in the projective plane $\mathcal{P}$ then all lines through $l \cap m$ are translation lines.*

Proof. Suppose $p = l \cap m$. Then p is a translation point in $\mathcal{P}$. Let l' be a line through p distinct from l and m. Since $\mathcal{P}$ is (p,m)-transitive, there is an elation with centre p and axis m mapping l to l'. (Why?) As l is a translation line, it follows that l' must be one too. □

It follows from this lemma that if there are three non-concurrent translation lines then all lines are translation lines. A plane with this property is called a *Moufang plane*. We have the following deep results, with no geometric proofs known.

4.5.2 Theorem. *If a projective plane has two translation lines, it is Moufang.* □

4.5.3 Theorem. *A finite Moufang plane is Desarguesian.* □

These are both proved in Chapter VI of Hughes and Piper[]. A Moufang plane which is not desarguesian can be constructed using the Cayley numbers. These form a vector space $\mathcal{O}$ of dimension eight over $\mathbb{R}$ with a multiplication such that

(a) if x and y lie in $\mathcal{O}$ and $xy = 0$ then either $x = 0$ or $y = 0$

(b) if x, y and z belong to $\mathcal{O}$ then $x(y+z) = xy + xz$ and $(y+z)x = yx + zx$.

It is worth noting that this multiplication is neither commutative, nor associative. To each element a of $\mathcal{O}$ we can associate an element ρ_a of $GL(\mathcal{O})$, defined by

$$\rho_a(x) = xa$$

for all x in $\mathcal{O}$. (This mapping is not a homomorphism.) Then ρ_a is injective and, since $\mathcal{O}$ is finite dimensional, it must be invertible. Moreover, if a and b both belong to $\mathcal{O}$ then $(\rho_a - \rho_b)x = xa - xb = x(a - b)$ and so $\rho_a - \rho_b$ is also invertible. Thus the set

$$\Sigma = \{\rho_x : x \in \mathcal{O} \setminus 0\}$$

gives rise to a spread of $\mathcal{O} \oplus \mathcal{O}$. As Σ is not closed under multiplication, the plane $\mathcal{P}$ determined by Σ cannot be Desarguesian. Since $\rho_{(x+y)} = \rho_x + \rho_y$ we see that Σ is a vector space over $\mathbb{R}$. By Lemma 2.3, this implies that $\mathcal{P}$ has two translation lines and hence that it is Moufang.

Chapter 5

Varieties

This chapter will provide an introduction to some elementary results in Algebraic Geometry.

5.1 Definitions

Let $V = V(n, \mathbb{F})$ be the n-dimensional vector space over the field $\mathbb{F}$. An *affine hypersurface* in V is the solution set of the equation $p(\mathbf{x}) = 0$, where p is a polynomial in n variables, together with the polynomial p. If $n = 2$ then a hypersurface is usually called a curve, and in three dimensions is known as a surface. An *affine variety* is the solution set of a set of polynomials in n variables together with the ideal, in the ring of all polynomials over $\mathbb{F}$, generated by the polynomials associated to the hypersurfaces. (This ideal is the ideal of polynomials which vanish at all points on the variety.) It is an important result that every affine variety can be realised as the solution set of a finite collection of polynomials. Affine varieties may also be defined as the intersection of a set of hypersurfaces.

A *projective hypersurface* is defined by a homogeneous polynomial in $n+1$ variables, usually $x_0, \ldots, x_n$. If p is such a polynomial and $p(\mathbf{x}) = 0$ then $p(\alpha\mathbf{x}) = 0$ for all scalars α in $\mathbb{F}$. The 1-dimensional subspaces spanned by the vectors $\mathbf{x}$ such that $p(\mathbf{x}) = 0$ are a subset of $PG(n, \mathbb{F})$, this subset is the projective hypersurface determined by p. A projective variety is defined in analogy to an affine variety. The 'ideal' of all homogeneous polynomials which vanish on the intersection is used in place of the ideal of all polynomials.

A *quadric* is a hypersurface defined by a polynomial of degree two. It may be affine or projective. A projective curve is a hypersurface in $PG(2, \mathbb{F})$ and a projective surface is a hypersurface in $PG(3, \mathbb{F})$. The hypersurface determined by the equation $g(\mathbf{x}) = 0$ will be denoted by $\mathcal{V}_g$. Only the context will determine if g is homogeneous or not. A *conic* is a quadric given by a polynomial of degree two.

Every affine variety gives rise to a projective variety in a natural way. This happens as follows. Let p be a polynomial in n variables $x_1, \dots, x_n$ with degree k. Let x_0 be a new variable and let q be the polynomial $x_0^k p(x_1/x_0, \dots, x_n/x_0)$. This a homogeneous polynomial of degree k in $n+1$ variables. By way of example, if p is the polynomial $x^2 - y - 1$ then q can be taken to be $x^2 - yz - z^2$. If we set $z = 1$ in q then we recover the polynomial p. Geometrically, this corresponds to deleting the line $z = 0$ from $PG(2, \mathbb{F})$ to produce an affine space. The only point on the curve $q(\mathbf{x}) = 0$ in $PG(2, \mathbb{F})$ and on the line $z = 0$ is spanned by $(0, 1, 0)^T$. The remaining points are spanned by the vectors $(x, x^2 - 1, 1)^T$, and these correspond to the points on the affine curve $p(\mathbf{x}) = 0$. We can also obtain affine planes by deleting lines other than $z = 0$. Thus if we delete the line $y = 0$ then remaining points on our curve are spanned by the vectors $(x, 1, z)^T$ such that $x^2 - z - z^2 = 0$. Although the original affine curve $x^2 - y - 1$ was a parabola, this curve is a hyperbola. This shows that each projective variety determines a collection of affine varieties. These affine varieties are said to be obtained by *dehomogenisation*. (But we will say this as little as possible.) Two affine varieties obtained in this way are called *projectively equivalent*. The number of different affine varieties that can be obtained from a given projective variety is essentially the number of ways in which it is met by a projective hyperplane.

The affine variety determined by a homogeneous polynomial g is said to be a *cone*. More generally, $\mathcal{V}_g$ is a cone at a point $\mathbf{a}$ if $g(\mathbf{y})$ is a homogeneous polynomial in $\mathbf{y} = \mathbf{x} - \mathbf{a}$. The projective variety associated with $\mathcal{V}_g$ is also said to be a cone at $\mathbf{a}$. Everything we have said so far is true whether or not the underlying field is finite or not. One difficulty in dealing with the finite case is that it may not be clear how many points lie on a given hypersurface. (Of course similar problems arise if we are working over the reals.) The next result is useful; it is known as Warning's theorem.

5.1.1 Theorem. *Let f be a polynomial of degree k in n variables over the field $\mathbb{F}$ with $q = p^r$ elements. If $k < n$ then the number of solutions of $f(\mathbf{x}) = 0$ is zero modulo p.*

Proof. We begin with some observations concerning $\mathbb{F}$. If $a \in \mathbb{F}$ then a^{q-1} is zero if $a = 0$ and is otherwise equal to 1. For a non-zero element λ of $\mathbb{F}$, consider the sum

$$S(\lambda) = \sum_{a \in \mathbb{F}} (\lambda a)^d.$$

Then $S(\lambda) = \lambda^d S(1)$. On the other hand, when a ranges over the elements of $\mathbb{F}$, so does λa. Hence $S(\lambda) = S(1)$, which implies that either $S(1) = 0$ or $\lambda^d = 1$. We may choose λ to be a primitive element of $\mathbb{F}$, in which case $\lambda^d = 1$ if and only if $q - 1$ divides d. This shows that if $q - 1$ does not divide d then $S(1) = 0$. If $q - 1$ divides d then $S(1) \equiv q - 1$ modulo p.

We now prove the theorem. The number of (affine) points $\mathbf{x}$ such that $f(\mathbf{x}) \neq 0$ is congruent modulo p to

$$\sum_{\mathbf{a} \in \mathbb{F}^n} f(\mathbf{a})^{q-1}. \tag{5.1}$$

The expansion of $f(\mathbf{x})^{q-1}$ is a linear combination of monomials of the form

$$x_1^{k(1)} \cdots x_n^{k(n)}$$

where

$$\sum_i k(i) \leq (q-1)k < (q-1)n.$$

This shows that for some i we must have $k(i) < q-1$. Therefore

$$\sum_{a_i \in \mathbb{F}} a_i^{k(i)}$$

is congruent to zero modulo p. This implies in turn that (5.1) is congruent to zero modulo p. □

The following result is due to Chevalley.

5.1.2 Corollary. *If f is a homogeneous polynomial of degree k in $n+1$ variables over the field F and $k \leq n$ then $\mathcal{V}_f$ contains at least one point of $PG(n, \mathbb{F})$.* □

These results generalise to sets of polynomials in n variables, subject to the condition that the sum of the degrees of the polynomials in the set is less than n. (See the exercises.)

5.2 The Tangent Space

Let f be a polynomial over $\mathbb{F}$ in the variables $x_0, \ldots, x_n$. By f_i we denote the partial derivative of f with respect to x_i. Even when $\mathbb{F}$ is finite, differentiation works more or less as usual. In particular both the product and chain rules still hold. The chief surprise is the constant functions are no longer the only functions with derivative zero. Thus, over $GF(2)$ we find that $\frac{\partial}{\partial x_i} x_i^2 = 0$. If f is homogeneous and $\mathbf{a} \in \mathcal{V}_f$ then the *tangent space of $\mathcal{V}_f$ at a* is the subspace given by the equation

$$\sum_{i=0}^{n} f_i(\mathbf{a}) x_i = 0.$$

It will be denoted by $T_{\mathbf{a}}(\mathcal{V}_f)$, or $T_{\mathbf{a}}(f)$. The tangent space at $\mathbf{a}$ always contains $\mathbf{a}$. This follows from Euler's Theorem, which asserts that if f is a homogeneous polynomial of degree k then

$$\sum_{i=0}^{n} x_i f_i = kf.$$

(The proof of this is left as a simple exercise. Note that it is enough to verify it for monomials.) The tangent space at the point $\mathbf{a}$ in the variety $\mathcal{V}$ defined by a set S of polynomials is defined to be the intersection of the tangent spaces of the hypersurfaces determined by the elements of S. If $f_i(\mathbf{a}) = 0$ for all i then $\mathbf{a}$ is a *singular point* of the hypersurface $\mathcal{V}_f$. When $\mathbf{a}$ is a singular point, $T_\mathbf{a}$ is the entire projective space and has dimension n. If $\mathbf{a}$ is not a singular point then $T_\mathbf{a}$ has dimension $n-1$ as a vector space. A singular point of a general variety can be defined as a point where the dimension of the tangent space is 'too large', but we will not go into details. A point which is not singular is called *smooth*, and a variety on which all points are non-singular is itself called *smooth* or *non-singular*. Questions about the behaviour of a variety at a particular point can usually be answered by working in affine space, since we can choose some hyperplane not on the point as the hyperplane at infinity.

5.3 Tangent Lines

If f is a homogeneous polynomial then the *degree* of the hypersurface $\mathcal{V}_f$ is the degree of f. The degree is important because it is an upper bound on the number of points in which $\mathcal{V}$ is met by a line. To see this, we proceed as follows. Assume that f is homogeneous of degree k, that a is a point and that b is a point not on $\mathcal{V}_f$. We consider the number of points in which $a \vee b$ meets $\mathcal{V}$. Suppose that $\mathbf{a}$ and $\mathbf{b}$ are vectors representing a and b. All points on $a \vee b$ are represented by vectors of the form $\lambda\mathbf{a} + \mu\mathbf{b}$. Thus the points of intersection of $a \vee b$ with $\mathcal{V}$ are determined by the values of λ and μ such that $f(\lambda\mathbf{a}+\mu\mathbf{b}) = 0$. Since $f(\mathbf{b}) \neq 0$ and f is homogeneous, all the points of intersection may be written in the form $\mathbf{a} + t\mathbf{b}$. Thus the number of points of intersection is the number of distinct solutions of

$$f(\mathbf{a}+t\mathbf{b}) = 0.$$

Now $f(\mathbf{a}+t\mathbf{b})$ is a polynomial of degree k in t, and hence has at most k distinct zeros. If the field we are working over is infinite then it can be shown that the degree of a hypersurface is actually equal to the maximum number of points in which it is met by a line. With finite fields this is not guaranteed—in fact the hypersurface itself is not guaranteed to have k distinct points on it. There is more to be said about the way in which a line can meet a hypersurface. Continuing with the notation used above, we can write

$$f(\mathbf{a}+t\mathbf{b}) = F^{(0)}(\mathbf{b}) + tF^{(1)}(\mathbf{b}) + \cdots + t^k F^{(k)}(\mathbf{b}), \tag{5.2}$$

where $F^{(i)}$ is a polynomial in the entries of $\mathbf{b}$, with coefficients depending on $\mathbf{a}$.. If the first nonzero term in (5.2) has degree m in t, we say that the *intersection multiplicity at a* of $a \vee b$ and $\mathcal{V}_f$ is m. If $a \in \mathcal{V}$ then $F^{(0)}(\mathbf{b}) = 0$; thus the intersection multiplicity is greater than zero if and only if a is on $\mathcal{V}$. We have

$$F^{(1)}(\mathbf{b}) = \sum_{i=0}^{n} f_i(\mathbf{a})\, b_i.$$

and so the intersection multiplicity is greater than 1 if and only if b lies in the tangent space $T_a(f)$. Since $a \in T_a(f)$, the point b is in $T_a(f)$ if and only if the line $a \vee b$ lies in $T_a(f)$. A line having intersection multiplicity greater than one with $\mathcal{V}_f$ is a *tangent line*. We have just shown that $T_a(f)$ is the union of all the tangent lines to $\mathcal{V}_f$ at a. A subspace is *tangent* to $\mathcal{V}_f$ at a if it is contained in $T_a(f)$. It is possible for the hypersurface $\mathcal{V}$ to completely contain a given line $a \vee b$. In this case the left side of (5.2) must be zero for all t, whence it follows that $a \vee b$ is a tangent. More generally, a subspace contained in $\mathcal{V}_f$ is tangent to $\mathcal{V}_f$ at each point in it. There is another important consequence of (5.2) which must be remarked on.

5.3.1 Lemma. *Any line meets a projective hypersurface of degree k in at most k points, or is contained in it.*

Proof. Let l be a line a let a be a point on l which is not on the hypersurface $\mathcal{V}_f$. The points of l on $\mathcal{V}$ are given by the solutions of (5.2). If this is identically zero then l is contained in the hypersurface, otherwise it is polynomial of degree k and has at most k zeros. □

We have only considered tangent spaces to hypersurfaces. Everything extends nicely to the case of varieties; we simply define the tangent space of the variety $\mathcal{V}$ at a to be the intersection of the tangent spaces at a of the hypersurfaces which intersect to form $\mathcal{V}$. Since we will not be working with tangent spaces to anything other than hypersurfaces, we say no more on this topic. The tangent space to a quadric is easily described, using the following result, which is a special case of Taylor's theorem.

5.3.2 Lemma. *Let f be a homogeneous polynomial of degree two in $n+1$ variables over the field $\mathbb{F}$ and let $H = H(f)$ be the $(n+1) \times (n+1)$ matrix with ij-entry equal to $\frac{\partial^2}{\partial x_i \partial x_j} f$. Then*

$$f(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda^2 f(\mathbf{x}) + \lambda\mu \mathbf{x}^T H \mathbf{y} + \mu^2 f(\mathbf{y}). \qquad \square$$

The matrix $H(f)$ is the *Hessian* of f. It is a symmetric matrix and, if the characteristic of $\mathbb{F}$ is even, its diagonal entries are zero. The tangent plane at the point $\mathbf{a}$ has equation $\mathbf{a}^T H \mathbf{x} = 0$, and therefore $\mathbf{a}$ is singular if and only $\mathbf{a}^T H = 0$. Consequently the quadric determined by f is smooth if $H(f)$ is non-singular. (However it is possible for the quadric to be smooth when H is singular. For example, consider any smooth conic in a projective plane over a field of even order.) If the characteristic of $\mathbb{F}$ is not even then $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H(f)\mathbf{x}$. Since we do not wish to restrict the characteristic of our fields, we will not be making use of this observation. One important consequence of Lemma 3.2 is that if a tangent to a quadric at a meets it at a second point b then it is contained in the quadric. (For these conditions imply that $f(\mathbf{a}) = \mathbf{a}^T H \mathbf{b} = f(\mathbf{b}) = 0$.) Since all lines through a singular point are tangents, it follows that a line which passes through a singular point and one other point must be contained in the quadric. Of course any line meeting a quadric in three or more points must be contained in it, by Lemma 3.1. A line which meets a quadric in two points is a *secant*.

5.3.3 Lemma. *Any line which meets a quadric in exactly one point is a tangent.*

Proof. Suppose l is a line passing through the point a on a given quadric $f(\mathbf{x}) = 0$ and that $b \in l$. Then the points of l on the quadric are given by the solutions of the quadratic in λ and μ:

$$\lambda^2 f(\mathbf{a}) + \lambda\mu \mathbf{a}^T A\mathbf{b} + \mu^2 f(\mathbf{b}) = 0.$$

Since $f(\mathbf{a}) = 0$ this quadratic has only one solution if and only $\mathbf{a}^T A\mathbf{b} = 0$, i.e., $\mathbf{b} \in T_a(f)$. Thus any line which meets the quadric in only the single point a must lie in the tangent space T_a. □

5.4 Intersections of Hyperplanes and Hypersurfaces

Suppose that f is a homogeneous polynomial defined over a field $\mathbb{F}$. Then f is irreducible if it does not factor over $\mathbb{F}$, and it is *absolutely irreducible* if it does not factor over the algebraic closure of $\overline{\mathbb{F}}$ of $\mathbb{F}$. If g is a factor of F over $\overline{\mathbb{F}}$ then $\mathcal{V}_g$ is a *component* of $\mathcal{V}_f$. Thus $\mathcal{V}_f$ is a union of components, although not necessarily a disjoint union. Over finite fields the situation is a little delicate, in that $\mathcal{V}_g$ may be empty. However this possibility will not be the source of problems—such components tend to remain completely invisible.

A hyperplane can be viewed as a projective space in its own right. By changing coordinates if needed, we may assume that the hyperplane has equation $x_0 = 0$. Suppose that f is homogeneous in $n + 1$ variables with degree k and that g is the polynomial obtained by setting x_0 equal to zero. Now g might be identically zero, in which case we must have $f = x_0 f'$ with f' a homogeneous polynomial of degree $k - 1$. Thus the hyperplane is a component of $\mathcal{V}_f$. If g is not zero then it is a homogeneous polynomial of degree k in n variables, and defines a nontrivial hypersurface. One interesting case is when the intersecting hyperplane is the tangent space to $\mathcal{V}_f$ at the point a. Every line through a in $T_a(f)$ is a tangent line to $\mathcal{V}_f$ and hence to $T_a(f) \cap \mathcal{V}_f$. Thus a is a singular point in the intersection. We will not have much cause to consider the intersection of two general hypersurfaces. There is one case concerning intersecting 'hypersurfaces' in projective planes where we will need some information. This result is called *Bézout's theorem.*

5.4.1 Theorem. *Let f and g be homogeneous polynomials over $\mathbb{F}$ in three variables with degree k and l respectively. Then either the curves $\mathcal{V}_f$ and $\mathcal{V}_g$ meet in at most kl points in $PG(2, \mathbb{F})$, or they have a common component.* □

In general two hypersurfaces of degrees k and l meet in a variety of degree kl. The theory describing the intersection of varieties is very complicated, even by the standards of Algebraic Geometry. The proof of the above result is quite simple though. (It can be found in "Algebraic Curves" by Robert J. Walker, Springer (New York) 1978. The proof of Bézout's theorem given there is over the complex numbers, but is valid for algebraically closed fields of any characteristic.) In making use of Bézout's lemma, we will need the following result, which is an extension of the fact that if a polynomial in one variable t over $\mathbb{F}$ vanishes at λ then it must have $t - \lambda$ as a factor.

5.4.2 Lemma. *Let f and g be polynomials in $n + 1$ variables over an algebraically closed field, with f absolutely irreducible. If $g(\mathbf{x}) = 0$ whenever $f(\mathbf{x}) = 0$ then f divides g.* □

As an immediate application of the previous ideas, we prove the following.

5.4.3 Lemma. *There is a unique conic through any set of five points which contains a 4-arc.*

Proof. Suppose that $abcd$ is a 4-arc. Let f be the homogeneous quadratic polynomial describing the conic formed by the union of the two lines $a \vee b$ and $c \vee d$, and let g be the quadratic describing the union of the lines $a \vee d$ and $b \vee c$. Consider the set of all quadratic polynomials of the form

$$\lambda f + \mu g. \tag{5.3}$$

Each of these is a quadratic, and thus describes a conic. If x is a point not on the 4-arc then the member of (5.3) with $\lambda = g(\mathbf{x})$ and $\mu = -f(\mathbf{x})$ vanishes at x and at each point of the 4-arc. This establishes the existence of a conic through any set five points containing a 4-arc. Suppose now that $\mathcal{C}$ and $\mathcal{C}'$ are two conics meeting on the 4-arc $abcd$ and the fifth point p. By Bézout's lemma, these two conics must have a common component. If the conics are distinct, this component must be described by a linear polynomial, i.e., it must be a line ℓ. Hence $\mathcal{C}$ and $\mathcal{C}'$ must each be the union of two lines, possibly the same line twice. But now each conic contains ℓ and at least two points from the 4-arc not on ℓ. We conclude that the conics must coincide. □

The hypersurfaces determined by the set of polynomials

$$\lambda f + \mu g, \quad \lambda, \mu \in \mathbb{F}$$

are said to form a *pencil*. We shall see that pencils can be very useful.

Chapter 6

Conics

We now begin our study of quadrics in $PG(2,\mathbb{F})$, i.e., conics. We will prove the well known theorems of Pappus and Pascal, along with Segre's theorem, which asserts that a $(q+1)$-arc in a projective plane over a field of odd order is a conic.

6.1 The Kinds of Conics

By Corollary 4.1.2, every conic over the field $\mathbb{F}$ contains at least one point. We will see that conics with only one point on them exist, but there is little to be said about them. There are two obvious classes of singular conics. The first consists of the ones with equations $(\mathbf{a}^T\mathbf{x})^2 = 0$, with all points singular. We will call this a *double line*. The second have equations $(\mathbf{a}^T\mathbf{x})(\mathbf{b}^T\mathbf{x}) = 0$, with $\mathbf{a}$ and $\mathbf{b}$ independent. The variety defined by such an equation is the union of two distinct lines; the point of intersection of these two lines is the unique singular point. A single point is also a conic. To see this, take an irreducible quadratic $f(x_0, x_1)$, then view it as a polynomial in three variables x_0, x_1 and x_2. Its solution set in the projective plane is the point $(0,0,1)^T$. Smooth conics do exist—the points of the form $(1,t,t^2)^T$ where t ranges over the elements of $\mathbb{F}$, together with the point $(0,0,1)^T$ provide one example. (This is the variety defined by the equation $x_0x_2 - x_1^2 = 0$. You should verify that it is smooth.) The four examples just listed exhaust the possibilities.

6.1.1 Theorem. *A conic in $PG(2,\mathbb{F})$ is either*

(a) a single point,

(b) a double line,

(c) the union of two distinct lines, or

(d) smooth, and a $(q+1)$-arc if $\mathbb{F}$ is finite with order q.

Proof. To begin we establish an important preliminary result, namely that if is a is a non-singular point in a conic $\mathcal{C} = \mathcal{V}_f$ then

$$|\mathcal{C}| = q + |T_a(\mathcal{C}) \cap \mathcal{C}|$$

(This implies that the cardinality of $\mathcal{C}$ is either $q+1$ or $2q+1$.) Suppose f is homogeneous of degree two and that $f(\mathbf{a}) = 0$. Then

$$f(\lambda\mathbf{a} + \mu\mathbf{x}) = \lambda\mu\mathbf{a}^T A\mathbf{x} + \mu^2 f(\mathbf{x}).$$

If $\mathbf{a}^T A\mathbf{x} \neq 0$, this implies that $f(\mathbf{x})\mathbf{a} - (\mathbf{a}^T A\mathbf{x})\mathbf{x}$ is a second point on the line through a and x which is on the conic. This shows that there is a bijection between the lines through a not in $T_a(f)$ and the points of $\mathcal{V}_f \setminus T_a(f)$. If a is a non-singular point then T_a is a line. By the previous lemma it contains either 1 or $q+1$ points of the conic. There are $q+1$ lines through any point in $PG(2,\mathbb{F})$. Thus if a is non-singular then the conic contains either $q+1$ or $2q+1$ points according as the tangent at a is contained in $\mathcal{C} = \mathcal{V}_f$ or not.

We now prove the theorem. Suppose that $\mathcal{C}$ is a conic. Assume first that it contains two singular points a and b. From our observations at the end of the previous section, all points on $a \vee b$ must belong to $\mathcal{C}$. If c is point of the conic not on $a \vee b$ then all points on $c \vee a$ and $c \vee b$ must also lie in $\mathcal{C}$. If x is a point in $PG(2,\mathbb{F})$ then there is a line through x meeting $c \vee a$, $c \vee b$ and $a \vee b$ in distinct points. Hence this line lies in $\mathcal{C}$ and so $x \in \mathcal{C}$. This proves that $\mathcal{C}$ is the entire plane, which is impossible. Thus we have shown that if $\mathcal{C}$ contains two singular points then it must consist of all points on the line joining them, i.e., it is a repeated line.

Assume then that $\mathcal{C}$ contains exactly one singular point, a say, and a further point b. Then $a \vee b$ is contained in $\mathcal{C}$. As there is only one singular point, there must a point of $\mathcal{C}$ which is not on $a \vee b$. The line joining this point to a is also in $\mathcal{C}$. This accounts for $2q+1$ points of $\mathcal{C}$, hence our conic must be the union of two distinct lines. Finally suppose that $\mathcal{C}$ contains at least two points, and no singular points. If $|\mathcal{C}| = 2q+1$ then each point of $\mathcal{C}$ must lie in a line contained in $\mathcal{C}$. Hence $\mathcal{C}$ must contain two distinct lines, and their point of intersection is singular. Consequently $\mathcal{C}$ can contain no lines, but must rather be a $(q+1)$-arc. □

This theorem is still valid over infinite fields, but the proof in this case is left to the reader. One consequence of it is that a conic is smooth if and only if it contains a 5-arc. In combination with Lemma 5.1, this implies that there is a unique smooth conic containing a given 5-arc.

6.2 Pascal and Pappus

The theorems of Pascal and Pappus are two of the most important results concerning projective planes over fields. We will prove both of these results using Bézout's lemma, and then give some of their applications. There are a few matters to settle before we can begin. A *hexagon* in a projective plane consists of cyclically ordered set of six points $A_0, A_1 \dots, A_5$, together with the six lines A_iA_{i+1}. Here the addition in the subscripts is computed modulo six. The six lines, which we require to be distinct, are the *sides* of the hexagon. Two sides are *opposite* if they are of the form A_iA_{i+1} and $A_{i+3}A_{i+4}$. Let $\mathbf{a}_{i,i+1}$, $i = 0, \dots, 5$ be the homogeneous coordinate vectors of the sides of the hexagon. Then the polynomial

$$f(\mathbf{x}) = (\mathbf{x}^T\mathbf{a}_{01})(\mathbf{x}^T\mathbf{a}_{23})(\mathbf{x}^T\mathbf{a}_{45}) \qquad (6.1)$$

is homogeneous with degree three. Similarly, the three sides opposite to those used in (6.1) determine a second cubic, g say. By Bézout's lemma, two cubics with no common component meet in at most nine points. A common component of our two cubics would have to contain a line, and our hypothesis that the sides are distinct prevents this. Therefore $\mathcal{V}_f$ and $\mathcal{V}_g$ meet in the six points of our hexagon, together with the points of intersection of the three pairs of opposite sides.

6.2.1 Theorem. *(Pascal). The six points of a hexagon lie on a conic if and only if the points of intersection of the three pairs of opposite sides lie on a line.*

Proof. Let $A_0, A_1 \dots, A_5$ be a hexagon. Suppose that the three points

$$A_0A_1 \cap A_3A_4,\ A_1A_2 \cap A_4A_5,\ A_2A_3 \cap A_0A_5$$

lie on a line l, with equation $\mathbf{a}^T\mathbf{x} = 0$. Let f and g be the two cubics defined above. For any scalars λ and μ, the polynomial $F = \lambda f + \mu g$ is cubic and contains the nine points in which $\mathcal{V}_f$ and $\mathcal{V}_g$ intersect. We wish to choose the scalars so that the line l is contained in $\mathcal{V}_F$. If l has only three points, there is no work to be done. Thus we may choose a fourth point p on l, and choose λ and μ so that $F(p) = 0$. Thus the cubic curve $\mathcal{V}_F$ meets the line l in four points, and if we extend $\mathbb{F}$ to its algebraic closure, then the line extending l still meets the extension of $\mathcal{V}_F$ in at least four points. Bézout's theorem now implies that l must be contained in the curve and so we deduce, by Lemma 4.4.2, that $F = (\mathbf{a}^T\mathbf{x})G$ for some polynomial F_1. But G must be homogeneous of degree two and therefore $\mathcal{V}_G$ is a conic. Thus $\mathcal{V}_F$ is the union of the line l and the conic $\mathcal{V}_G$. If the hexagon is contained in the union of two lines then it is on a conic, and we are finished. Otherwise a simple check shows that no points on the hexagon lie on L (do it), hence they line on the conic. This proves the first part of the theorem.

Assume now that the points of the hexagon lie on a conic. There is no loss on assuming that this conic is not a double line or a single point. Thus it is either the union of two distinct lines, or is smooth. It is convenient to treat these two cases separately. Supose then that our conic is the union of the two lines l and m, with respective equations $\mathbf{a}^T\mathbf{x} = 0$ and $\mathbf{b}^T\mathbf{x} = 0$. As the sides of our hexagon are distinct, no four points of it are collinear. (Why?) Hence three points of the hexagon lie on l and three on m. In particular, $p = l \cap m$ is not a point of the hexagon. Now choose λ and μ so that $F = \lambda f + \mu g$ passes through

p. Then the lines l and m each meet the cubic F in four points, and so they must lie in $\mathcal{V}_F$. Hence F is divisible by $(\mathbf{a}^T\mathbf{x})(\mathbf{b}^T\mathbf{x})$ and the quotient with respect to this product must be linear. Thus F is the union of three lines. Consequently the points of intersection of the opposite sides of the hexagon must be collinear.

There remains the case that the points of the hexagon lie on a smooth conic $\mathcal{C}$, with equation $h(\mathbf{x}) = 0$. This conic meets any curve of the form

$$F(\mathbf{x}) := \lambda f(\mathbf{x}) + \mu g(\mathbf{x}) = 0 \tag{6.2}$$

in at least the six points of the hexagon. As $|\mathcal{C}| \geq 6$, our field must have order at least five. If it is exactly five then $\mathcal{C}$ is contained in the solution set of (6.2) for any choice of scalars; otherwise we may choose a point p of $\mathcal{C}$ not in the hexagon and then choose λ and μ so that $\mathcal{V}_F$ meets $\mathcal{C}$ in at least seven points. By Bézout's theorem, this implies that these two curves have a common component. The only component of $\mathcal{C}$ is $\mathcal{C}$ itself, thus $F = hG$ for some linear polynomial G. Hence $\mathcal{V}_F$ is the union of a line and the conic $\mathcal{C}$, and the points of intersection of the opposite sides of our hexagon must be on the line. □

Pappus' theorem is the assertion that the intersections of the opposite sides of a hexagon are collinear if the points of the hexagon lie on two lines. It is particularly important because it can be proved that a projective plane has the form $PG(2, \mathbb{F})$, where $\mathbb{F}$ is a field, if and only if Pappus' theorem holds. Thus, if we could prove geometrically that Pappus' theorem held in all finite Desarguesian planes then we would have a geometric proof that a finite skew field is a field. No such proof is known. Planes for which Pappus' theorem is valid are called *Pappian*. All Pappian planes are, of course, Desarguesian.

6.3 Automorphisms of Conics

If $\mathcal{C}$ is a conic described by the equation $f(\mathbf{x}) = 0$ and $\tau \in PGL(3, \mathbb{F})$ then we let f^τ denote the polynomial defining the conic $\mathcal{C}\tau$. The *automorphism group* of a conic in the Pappian plane $PG(2, \mathbb{F})$ is the subgroup of $PGL(3, \mathbb{F})$ which fixes it as a set. The concept is well defined in all cases, but we will mainly be interested in automorphisms of smooth conics. Our previous theorem implies that smooth conics have many automorphisms.

6.3.1 Theorem. *Let abcd be a 4-arc in a Pappian projective plane and let $\mathcal{C}$ be a conic containing it. Then there is an involution τ in the automorphism group of $\mathcal{C}$ such that $a\tau = d$ and $b\tau = c$.*

Proof. As $PGL(3, \mathbb{F})$ is transitive on ordered 4-arcs, it contains an element τ mapping $abcd$ to $badc$. Hence τ fixes both the conics $ac \cup bd$ and $ab \cup cd$. Suppose that these conics are defined by the polynomials f and g repectively. For any λ and μ in $\mathbb{F}$, we find that

$$(\lambda f + \mu g)\tau = \lambda f^\tau + \mu g^\tau = \lambda f + \mu g.$$

Hence τ fixes each quadric in the pencil determined by f and g. Since every conic containing the given 4-arc belongs to this pencil, this proves the theorem. □

One immediate consequence of this theorem is the following result.

6.3.2 Corollary. *Let $\mathcal{C}$ be a smooth conic in a Pappian plane. Then its automorphism group acts sharply 3-transitively on the points in it.*

Proof. If $|\mathbb{F}| = 2$ or 3, this result can be verified easily. Assume then that $|\mathbb{F}| > 3$. From the theorem, $\mathrm{Aut}(\mathcal{C})$ is 2-transitive on the points of $\mathcal{C}$. To prove that $\mathrm{Aut}(\mathcal{C})$ is 3-transitive it will suffice to prove that if A, B, C and D are four points on $\mathcal{C}$ then there is an automorphism of it fixing A and B and mapping C to D.

Let X be a fifth point on the conic. By the theorem, there is an involution in $\mathrm{Aut}(\mathcal{C})$ swapping A and B, and sending C to X. Similarly, there is an involution swapping B and A and sending X to D. The product of these two involutions is the required automorphism. Suppose that A, B and C are three points on the conic. Any automorphism which fixes these three points must fix the tangents at A and B. Hence it fixes their point of intersection, which we denote by P. Thus the automorphism fixes each point in a 4-arc, and the only element of $PGL(3, \mathbb{F})$ which fixes a 4-arc is the identity. □

It follows at once from the corollary that if $|\mathbb{F}| = q$ then $|\mathrm{Aut}(\mathcal{C})| = \Pi^{\ni} - \Pi$. We have already seen that the conics in $PG(2, \mathbb{F})$ correspond to the points in $PG(5, \mathbb{F})$, and are thus easily counted, there are

$$[6] = q^5 + q^4 + q^3 + q^2 + q + 1$$

of them. As for the smooth conics, we have:

6.3.3 Lemma. *Let $\mathbb{F}$ be the field with q elements, where $q > 3$. Then the number of smooth conics in $PG(2, \mathbb{F})$ is equal to $q^5 - q^2$.*

Proof. Let n_k denote the number of ordered k-arcs and let N be the number of smooth conics. Then, as we noted at the end of Section 6, there is a unique smooth conic containing a given 5-arc. Hence

$$N(q + 1)q(q - 1)(q - 2)(q - 3) = n_5. \tag{6.3}$$

We find that

$$n_3 = (q^2 + q + 1)(q^2 + q)q^2.$$

Let ABC be a 3-arc. There $q - 1$ lines through A which do not pass through B or C, and on each of these lines there are $q - 1$ points which do not lie on any line joining B and C. Thus we can extend a ABC to a 4-arc using any one of $(q - 1)^2$ points, and so $n_4 = (q - 1)^2 n_3$. There are $q - 2$ lines through a point in a 4-arc $ABCD$ which do not meet a second point on the arc, and each of these lines contains $q - 3$ points not on the lines BC, BD or CD. Thus $n_5 = (q - 2)(q - 3)n_4$. Accordingly

$$n_5 = (q - 3)(q - 2)(q - 1)^2 q^3 (q + 1)(q^2 + q + 1)$$

and, on comparing this with (6.3), we obtain that $N = (q^2 + q + 1)q^2(q - 1)$ as claimed. □

The group $PGL(3,\mathbb{F})$ permutes the smooth conics in $PG(2,\mathbb{F})$ amongst themselves. The number of conics in the orbit containing $\mathcal{C}$ is equal to

$$|PGL(3,\mathbb{F})|/|\mathrm{Aut}(\mathcal{C})|.$$

The order of $PGL(3,\mathbb{F})$ is

$$(q-1)^{-1}(q^3-1)(q^3-q)(q^3-q^2) = (q^2+q+1)(q+1)q^3(q-1)^2.$$

Since the automorphism group of a smooth conic has order q^3-q, the orbit of $\mathcal{C}$ has cardinality equal to

$$(q^2+q+1)(q+1)^2q^3(q-1)^2/(q^3-q) = (q^5-q^2).$$

As there are altogether q^5-q^2 smooth conics, this implies the following.

6.3.4 Theorem. *All smooth conics in the Pappian plane $PG(2,\mathbb{F})$ are equivalent under the action of $PGL(3,\mathbb{F})$.* □

6.4 Ovals

An oval in a projective plane of order q, i.e., with $q+1$ points on each line, is simply a $(q+1)$-arc. Every smooth conic in a Pappian plane is a $(q+1)$-arc; we show now that ovals have many properties in common with conics. As usual, some definitions are needed. Let $\mathcal{K}$ be a k-arc. A *secant* to $\mathcal{K}$ is a line which meets it in two points, a *tangent* meets it in one point. A line which does not meet the arc is an *external line*. Since no line meets $\mathcal{K}$ in three points, it has exactly $\binom{k}{2}$ secants. Each point in $\mathcal{K}$ lies on $k-1$ of these secants, whence there are $q+2-k$ tangents through each point and $k(q+2-k)$ tangents altogether. An immediate consequence of these deliberations is that a k-arc has at most $q+2$ points on it. (If q is odd this bound can be reduced to $q+1$. Proving this is left as an exercise.) Our next result is an analog of the fact that a circle in the real plane divides the points into three classes:

(a) the points outside the circle, which each lie on two tangents

(b) the points on the circle, which lie on exactly one tangent

(c) the points inside the circle, which lie on no tangents.

6.4.1 Lemma. *Let $\mathbb{F}$ be the field of order q, where q is odd, and let $\mathcal{Q}$ be a $(q+1)$-arc in $PG(2,\mathbb{F})$. Then there are $\binom{q+1}{2}$ points, each lying on exactly two tangents to $\mathcal{Q}$, and $\binom{q}{2}$ points which lie on none.*

Proof. Suppose P is a point on a tangent to $\mathcal{Q}$, but not on $\mathcal{Q}$. Then the lines through P meet $\mathcal{Q}$ in at most two points, and thus they partition the points of $\mathcal{Q}$ into pairs and singletons. Each singleton determines a tangent to $\mathcal{Q}$ through

P. Since $q + 1$ is even, P lies on an even number of tangents. As P is on one tangent, it therefore lies on at least two. On the other hand, each pair of tangents to $\mathcal{Q}$ meet at a point off $\mathcal{Q}$, and this point is on two tangents. Thus there are at most $\binom{q+1}{2}$ triples formed from a pair of distinct tangents and their point of intersection. This implies that any point off $\mathcal{Q}$ which is on a tangent is on exactly two. □

When q is even, the tangents to a $(q+1)$-arc behave in an unexpected fashion.

6.4.2 Lemma. *Let $\mathbb{F}$ be the field of order q, where q is even, and let $\mathcal{Q}$ be a $(q+1)$-arc in $PG(2,\mathbb{F})$. Then the tangents to $\mathcal{Q}$ are concurrent. Thus there is one point which lies on all tangents to $\mathcal{Q}$, and the remaining points off $\mathcal{Q}$ all lie on exactly one tangent.*

Proof. Let P and Q be two distinct points on $\mathcal{Q}$. Since the number of points in the oval is odd, each point on the line PQ which is not on $\mathcal{Q}$ must lie on a tangent to it. As P and Q both lie on tangents, it follows that each point on PQ is on a tangent. The number of tangents to $\mathcal{Q}$ is $q + 1$ and the number of points on PQ is also $q + 1$. Thus each point on a secant to $\mathcal{Q}$ is on a unique tangent. Now let K be the point of intersection of two tangants which do not meet on $\mathcal{Q}$. Then K cannot lie on any secant, and so all lines through K are tangents to $\mathcal{Q}$. □

The point K is called the *nucleus* or *knot* of the oval. The oval, together with its nucleus forms a $(q + 2)$-arc. A $(q + 2)$-arc is sometimes called a *hyperoval*. Since we can delete any point from a hyperoval to obtain an oval, a given oval can thus be used to form a number of distinct ovals. In particular, if we start with a conic in a Pappian plane of even order, we can construct $(q + 1)$-arcs which are not conics.

6.5 Segre's Characterisation of Conics

B. Segre proved that, if q is odd, any $(q + 1)$-arc in the projective plane over $GF(q)$ is a conic. We now present the proof of this important result. We first describe one property of $(q+1)$-arcs in projective planes over fields of odd order.

6.5.1 Lemma. *Let $\mathcal{Q}$ be a $(q+1)$-arc in the projective plane over $GF(q)$, where q is odd. Let A_0, A_1 and A_2 be three distinct points on $\mathcal{Q}$ and let l_0, l_1 and l_2 be the tangents at these three points. Then the triangle $A_0A_1A_2$ is in perspective with the triangle formed by the points $l_1 \cap l_2$, $l_0 \cap l_2$ and $l_0 \cap l_1$.*

Proof. Since $PGL(3,\mathbb{F})$ is transitive on 3-arcs, we may take the points A_0, A_1 and A_2 to be represented by

$$(1,0,0)^T,\ (0,1,0)^T,\ (0,0,1)^T$$

respectively. The line A_0A_1 can be represented by $(0,0,1)$ and A_0A_2 by $(0,1,0)$. Hence the tangent to $\mathcal{Q}$ at A_0 can be represented by a vector of the form

$$(0,1,0) - k_0(0,0,1) = (0,1,-k_0)$$

Similarly the tangents at A_1 and A_2 are represented by $(-k_1,0,1)$ and $(1,-k_2,0)$. We will show that $k_0k_1k_2 = -1$, and then deduce the lemma from this. Let B be a point $(b_0,b_1,b_2)^T$ on $\mathcal{Q}$ distinct from A_0. The line A_0B can be taken to be represented by the vector

$$(0,1,-h_0)$$

for some scalar $h_0 = h_0(B)$, and this scalar is non-zero if the line does not pass through A_1. Similarly, if B is distinct from A_1, the line A_1B can be represented by

$$(-h_1,0,1)$$

and, if B is distinct from A_2 then A_2B can be represented by

$$(1,-h_2,0).$$

Since B lies on the lines represented by these three vectors, we find that

$$b_1 = h_0b_2,\ b_2 = h_1b_0,\ b_0 = h_2b_1.$$

This shows that if one coordinate of B is zero than all coordinates are zero. Hence none of b_0, b_1 and b_2 is equal to zero, and this implies in turn that

$$h_0(B)h_1(B)h_2(B) = 1. \tag{6.4}$$

Conversely, if h is an element of $\mathbb{F}$ different from 0 and k_0 then the line represented by $(0,1,-h)$ passes through A_0 and some point on $\mathcal{Q}$ distinct from A_1 and A_2. The product of the $q-1$ non-zero elements of $\mathbb{F}$ is equal to -1 and so the product of the parameters $h_0(B)$ as B ranges over the points of $\mathcal{Q}$ distinct from A_0, A_1 and A_2 must be $-1/k_0$. It follows now using (6.4) that

$$\frac{(-1)^3}{k_0k_1k_2} = 1$$

and hence $k_0k_1k_2 = -1$, as claimed. The tangents at A_0 and A_1 meet at $(1,k_0k_1,k_1)^T$ and the line joining this to A_2 is represented by the vector

$$(k_0k_1,-1,0).$$

The tangents at A_1 and A_2 meet at $(k_2,1,k_1k_2)$ and the line joining this to A_0 is given by

$$(0,k_1k_2,-1).$$

The tangents at A_2 and A_0 meet at $(k_0k_2,k_0,1)$ and this joined to A_1 by the line

$$(-1,0,k_0k_2).$$

These three lines are concurrent, passing through the point $(1,k_0k_1,-k_1)$. □

6.5.2 Theorem. *(Segre). If q is odd then any $(q+1)$-arc in the projective plane over $GF(q)$ is a conic.*

Proof. We continue with the notation of the previous lemma. Since $PGL(3,\mathbb{F})$ is transitive on ordered 4-arcs, we may assume without loss that the triangles of the lemma are in perspective from the point $(1,1,1)^T$, i.e., that $k_0 = k_1 = k_2 = -1$. Let B be a fourth point on $\mathcal{Q}$ with coordinate vector $(x_1, x_2, x_3)^T$ and tangent vector (l_0, l_1, l_2). The line joining B to the intersection of the tangents at A_0 and A_1 has coordinate vector of the form

$$\alpha_0(0,1,1) + \beta_0(1,0,1).$$

Since this line passes through B we may take $\alpha = x_0 + x_2$ and $\beta = -(x_1 + x_2)$. Similarly the line through A_0 and the intersection of the tangents at A_1 and B can be taken to have coordinate vector

$$l_0(1,0,1) - (l_0, l_1, l_2)$$

while the line joining A_1 to the intersection of the tangents at B and A_0 has coordinate vector

$$(l_0, l_1, l_2) - l_1(0,1,1).$$

Since these three coordinate vectors represent concurrent lines, they are linearly dependent. Since the tangents at A_0, A_1 and B are not concurrent, they are linearly independent. Since we have the former written as linear combinations of the latter, it follows that the matrix of coefficients

$$\begin{pmatrix} 0 & x_0 + x_2 & -(x_1 + x_2) \\ -1 & 0 & l_0 \\ 1 & -l_1 & 0 \end{pmatrix}$$

must have determinant zero. This implies that

$$l_1(x_1 + x_2) = l_0(x_0 + x_2).$$

The last identity was derived by working with the three points B, A_0 and A_1. If instead we use B, A_1 and A_2 we obtain

$$l_2(x_0 + x_2) = l_1(x_0 + x_1).$$

Thus the respective ratios between l_0, l_1 and l_2 are the same as the ratios between $x_1 + x_2$, $x_0 + x_2$ and $x_0 + x_1$. We also have

$$l_0 x_0 + l_1 x_1 + l_2 x_2 = 0,$$

since B lies on the tangent to $\mathcal{Q}$ at B. Hence we get

$$0 = (x_1 + x_2)x_0 + (x_0 + x_2)x_1 + (x_0 + x_1)x_2 = 2(x_0x_1 + x_1x_2 + x_0x_2).$$

Since our field has odd order, we can now divide this by two, and thus deduce that the points of $\mathcal{Q}$ distinct from A_0, A_1 and A_2 lie on the conic

$$x_0x_1 + x_1x_2 + x_0x_2 = 0.$$

It is trivial to check that A_0, A_1 and A_2 lie on it too. □

More general results are known. Every q-arc in a projective plane over a field of odd order $q \geq 5$ must be contained in a conic. (We present one proof of this in the next section. A more elementary proof will be found in Lüneburg.) In addition to its beauty, Segre's theorem has a number of important applications. We will meet some of these later. There do exist $(q+1)$-arcs in projective planes over fields of even order which are not related to conics. (See the Exercises for an example.)

6.6 q-Arcs

Let $\mathcal{K}$ be a k-arc in the projective plane over the field of order q. Then each point in the arc lies on

$$(q+1) - (k-1) = q+2-k$$

tangents to the arc. These tangents thus form a set of $k(q+2-k)$ points in the dual space. We have the following result. A proof will be found in Hirschfeld [PGOFF].

6.6.1 Theorem. *(Segre). Let $\mathcal{K}$ be a k-arc in the projective plane over the field of order q. Then the points in the dual plane corresponding to the tangents to the arc lie on a curve. This curve does not contain a point corresponding to a secant, and has degree $q+2-k$ if q is even and degree $2(q+2-k)$ if q is odd.* □

6.6.2 Corollary. *(Segre). Let $\mathcal{K}$ be a q-arc in the projective plane over the field with order q, and let q be odd. Then $\mathcal{K}$ is contained in a conic.*

Proof. We have already proved that every 3-arc in contained in a 4-arc, so we may assume that $q > 3$. By the theorem, there is a curve of degree four $\mathcal{C}$ in the dual plane which contains the $2q$ points corresponding to the tangents to $\mathcal{K}$, and none of the points corresponding to the secants. Let a be a point off $\mathcal{K}$. Since q is odd, the number of tangents to $\mathcal{K}$ through a is odd. Suppose that a lies on at least five tangents to $\mathcal{K}$. The lines through a correspond to the points on a line ℓ in the dual plane, and ℓ meets $\mathcal{C}$ in at least five points. Since $\mathcal{C}$ has degree four, Bézout's theorem yields that ℓ must be a component of $\mathcal{C}$. Thus all the points of ℓ are on $\mathcal{C}$, and so none of the lines through a can be secants to $\mathcal{K}$. Therefore all the lines through a which meet $\mathcal{K}$ are tangents, and so $\mathcal{K} \cup a$ is a $(q+1)$-arc. Since q is odd, all $(q+1)$-arcs are conics by Theorem 5.2.

We can complete the proof by showing that for any q-arc, there is a point a on at least five tangents. If $y \notin \mathcal{K}$, let t_y be the number of tangents to $\mathcal{K}$ through y. By counting the pairs (ℓ, y), where y is a point off $\mathcal{K}$ and ℓ is a tangent through y, we find that

$$\sum_{y \notin \mathcal{K}} t_y = 2q^2$$

and by counting the triples (ℓ, ℓ', y) where ℓ and ℓ' are distinct tangents and $y = \ell \cap \ell'$, we obtain

$$\sum_{y \notin \mathcal{K}} t_y(t_y - 1) = 2q(2q - 2).$$

Together these equations imply that

$$\sum_{y \notin \mathcal{K}} (t_y - 1)(t_y - 3) = (q - 1)(q - 3).$$

Since q is odd, t_y is odd for all points y not on $\mathcal{K}$. As $q > 3$, the last equation thus implies that $t_y \geq 5$ for some point y not on $\mathcal{K}$. □

The above proof is an improvement on the original argument of Segre, due to Thas.

Chapter 7

Polarities

In this chapter we study polarities of projective geometries. We will see that they are closely related to quadrics.

7.1 Absolute Points

A *polarity* of a symmetric design is a bijective mapping ϕ sending its points to its blocks and its blocks to its points, such that if $x \in y^\phi$ then $y \in x^\phi$. A point x such that $x \in x^\phi$ is called *absolute*, and if every point is absolute we say that ϕ is a *null polarity*. A polarity of a design determines automatically a polarity of the complementary design. (This will be null if and only if ϕ has no absolute points.) The points and hyperplanes of a projective geometry form a symmetric design. The mapping which takes the point with homogeneous coordinate vector $\mathbf{a}$ to the hyperplane with vector $\mathbf{a}^T$ is our first example of a polarity. Let $\mathcal{D}$ be a symmetric design with points $v_1, \dots, v_n$ and a polarity ϕ. Then the incidence matrix, with ij-entry equal to 1 if $x_i \in x_j^\phi$ and zero otherwise, is symmetric. (In fact, a symmetric design has a polarity if and only if it has a symmetric incidence matrix.)

7.1.1 Theorem. *Let $\mathcal{D}$ be a symmetric (v, k, λ)-design with a polarity ϕ. Then*

(a) if $k - \lambda$ is not a perfect square, ϕ has exactly k absolute points,

(b) if ϕ is null then $\sqrt{k-\lambda}$ is an integer and divides $v - k$,

(c) if ϕ has no absolute points then $\sqrt{k-\lambda}$ is an integer and divides k.

Proof. Let N be the incidence matrix of $\mathcal{D}$. As just noted, we may assume that N is symmetric, whence we have

$$N^2 = (k - \lambda)I + \lambda J. \tag{7.1}$$

(Here J is the matrix with every entry equal to 1.) The number of absolute points of the polarity is equal to $\operatorname{tr} N$, which is in turn equal to the sum of the eigenvalues of N. From (7.1) we see that the eigenvalues of N^2 coincide with the eigenvalues of $(k-\lambda)I+\lambda J$. This means that N^2 must have as its eigenvalues

$$k-\lambda+(v-1)\lambda$$

with multiplicity one and $k-\lambda$, with multiplicity $v-1$. A simple design theory calculation shows that $k-\lambda+(v-1)\lambda=k^2$. The eigenvalues of N^2 are the squares of the eigenvalues of N. As each row of N sums to k, we see that k is an eigenvalue of N. Since k^2 is a simple eigenvalue of N^2, it follows that $-k$ cannot be an eigenvalue of N. Hence N has $v-1$ eigenvalues equal to either $\sqrt{k-\lambda}$ or $-\sqrt{k-\lambda}$. Suppose that there are exactly a eigenvalues of the first kind and b of the second. Then

$$\operatorname{tr} N = k+(a-b)\sqrt{k-\lambda} \tag{7.2}$$

and, as $\operatorname{tr} N$, k, a and b are all integers, this implies that either $a=b$ or $(k-\lambda)$ is a perfect square. This proves (a) in the statement of the theorem. If the polarity is null then $\operatorname{tr} N = v$, whence (7.2) implies that

$$\sqrt{k-\lambda}=\frac{v-k}{b-a}.$$

Since the right hand side is rational this implies again that $k-\lambda$ is a perfect square, and in addition that $\sqrt{k-\lambda}$ must divide $v-k$. Finally, (c) follows from (b) applied to the complement of the design $\mathcal{D}$. □

7.1.2 Corollary. *Every polarity of a finite projective space has an absolute point.*

Proof. Continuing with the notation of the theorem, we see that if $k-\lambda$ is a perfect square then $\sqrt{k-\lambda}$ divides k if and only if it divides λ. For a projective geometry of rank n and order q we have

$$v=[n],\quad k=[n-1],\quad \lambda=[n-2],$$

whence $k-\lambda=q^{n-1}$ and $v-k=q^n$. Therefore k and λ are coprime for all possible values of q and n. □

7.2 Polarities of Projective Planes

The results in this section are valid for all projective planes, Desarguesian or not. If x is a point or line in a projective plane and ϕ is a polarity of the plane then we denote the image of x under ϕ by x^ϕ.

7.2.1 Lemma. *Let ϕ be a polarity of a projective plane. Then each absolute line contains exactly one absolute point, and each absolute point is on exactly one absolute line.*

Proof. The second statement is the dual of the first, which we prove as follows. Suppose a is an absolute point and that b is a second absolute point on $\ell = a^\phi$. Then $a \in b^\phi$ since $b \in a^\phi$. So

$$a \in \ell \cap b^\phi.$$

Now $b^\phi \neq \ell$, because $a^\phi = b^\phi$ implies $a = b$. Hence

$$a = \ell \cap b^\phi$$

Since $b = \ell \cap b^\phi$, this proves that $a = b$. □

7.2.2 Theorem. *Let ϕ be a polarity of a projective plane of order n. Then ϕ has at least $n + 1$ absolute points. These points are collinear if n is even and form a $q + 1$-arc otherwise.*

Proof. Let m be a non-absolute line. We show first that the number of absolute points on m is congruent to n, modulo 2. Suppose $a \in m$. If a is not an absolute point then $b = a^\phi \cap m$ is a point on m distinct from a. Further, b^ϕ contains both a and m^ϕ; hence it is a line through a distinct from m. Thus $b^\phi \cap m = a$, and we have shown that the pairs

$$\{a, a^\phi \cap m\}$$

partition the non-absolute points on m into pairs. This proves the claim.

Assume now that n is even and let p be a non-absolute point. The $n + 1$ lines through p partition the remaining points of the plane. As each line must contain an absolute point ($n+1$ is odd) there are at least $n+1$ absolute points. Suppose that there are exactly $n + 1$ absolute points, and let x and y be two of them. If there is a non-absolute point q on $x \vee y$ then the argument we have just shows that the n lines through q distinct from $x \vee y$ contain at least n distinct absolute points. Taken with x and y we thus obtain at least $n + 2$ absolute points. This completes the proof of the theorem when n is even.

Assume finally that n is odd and let p be an absolute point. Then p^ϕ is the unique absolute line through p and so there are n non-absolute lines through p. Each of these contains an even number of absolute points, and hence at least one absolute point in addition to p. This shows that there are at least $n + 1$ absolute points. If there are exactly $n + 1$, this argument shows that each line through p contains either one or two absolute points. As our choice of p was arbitrary, it follows that the absolute points form an arc. □

7.2.3 Theorem. *Let ϕ be a polarity of a projective plane of order n. Then ϕ has at most $n^{3/2}+1$ absolute points. If this bound is achieved then the absolute points and non-absolute lines form a 2-$(n^{3/2} + 1, n^{1/2} + 1, 1)$ design.*

Proof. Denote the number of absolute points by s and k_i be the number of absolute points on the i-th non-absolute line. (The ordering is up to you.) Let $N = n^2 + n + 1 - s$; thus N is the number of non-absolute lines. Consider the ordered pairs (p, ℓ) where p is a absolute point and ℓ is a non-absolute line on p. Each absolute point is on n non-absolute lines, so counting these pairs in two ways yields

$$ns = \sum_{i=1}^{N} k_i. \tag{7.3}$$

Next we consider the ordered triples (p, q, ℓ) where p and q are absolute points on the non-absolute line ℓ. Counting these in two ways we obtain

$$s(s-1) = \sum_{i=1}^{N} k_i(k_i - 1). \tag{7.4}$$

The function $x^2 - x$ is convex and so

$$\sum_{i=1}^{N} \frac{k_i(k_i-1)}{N} \geq \frac{\sum_{i=1}^{N} k_i}{N} \left(\frac{\sum_{i=1}^{N} k_i}{N} - 1 \right),$$

with equality if and only if the k_i are all equal. Using (**??**) and (7.4), this implies that $n^2 s \leq (s+n-1)N$. Recalling now that $N = n^2+n+1-s$ and indulging in some diligent rearranging, we deduce that $(s-1)^2 \leq n^3$, with equality holding if and only if the k_i are equal. This yields the theorem. □

A 2-$(m^3+1, m+1, 1)$-design is called a *unital*. We will see how to construct examples in the following sections. We record the following special properties of the set of absolute points of a polarity realising the bound of the theorem.

7.2.4 Lemma. *Let ϕ be a polarity of a projective plane of order n having $n^{3/2}+1$ fixed points. Then every line meets the set $\mathcal{U}$ of absolute points of ϕ in 1 or $n^{1/2}+1$ points. For each point u in $\mathcal{U}$ there is a unique line ℓ such that $\ell \cap \mathcal{U} = u$, and for each point v off $\mathcal{U}$ there exactly $n^{1/2}+1$ lines through it which meet $\mathcal{U}$ in one one point.* □

7.3 Polarities of Projective Spaces

We are now going to study polarities of projective spaces over fields, and will give a complete description of them. The key observation is that a polarity is a collineation from $PG(n, \mathbb{F})$ to its dual and is therefore, by the Fundamental Theorem of Projective Geometry, induced by a semi-linear mapping. Let ϕ be a polarity of $PG(n-1, \mathbb{F})$. Then there is an invertible $n \times n$ matrix A over $\mathbb{F}$ and a field automorphism τ such that, if a is represented by the vector $\mathbf{a}$ then A^{ϕ} is represented by $(\mathbf{a}^{\tau})^T A$. Thus a^{ϕ} is the hyperplane with equation $(\mathbf{a}^{\tau})^T A\mathbf{x} = 0$. Since ϕ is a polarity,

$$(\mathbf{x}^{\tau})^T A\mathbf{y} = 0 \iff (\mathbf{y}^{\tau})^T A\mathbf{x} = 0.$$

But $(\mathbf{y}^\tau)^T A\mathbf{x} = 0$ if and only if $\mathbf{x}^T A^T \mathbf{y}^\tau = 0$, and this is equivalent to requiring that $(\mathbf{x}^T A^T)^\tau \mathbf{y} = 0$. Hence $(\mathbf{x}^\tau)^T A$ and $(\mathbf{x}^T A^T)^{\tau^{-1}}$ are coordinate vectors for the same hyperplane. This implies that $A^T\mathbf{x}^\tau = \kappa_1 (A\mathbf{x})^{\tau^{-1}}$ for some non-zero scalar κ_1, and so

$$A^{-1}(A^\tau)^T \mathbf{x}^{\tau^2} = \kappa\mathbf{x} \tag{7.5}$$

with $\kappa = \kappa_1^\tau$.

Since $A^{-1}(A^\tau)^T$ is a linear and not a semilinear mapping, it follows from (7.5) that $\mathbf{x}^{\tau^2}$ must lie in $V(n,\mathbb{F})$, and hence that $\tau^2 = 1$. Therefore (7.5) implies that $A^{-1}(A^\tau)^T = \kappa I$ and so we have shown that every polarity is determined by a field automorphism τ of order dividing two and a linear mapping A such that $(A^\tau)^T = \kappa A$. Now

$$A = A^{\tau^2} = ((A^\tau)^T)^\tau)^T = ((\kappa A)^\tau)^T = \kappa^\tau (A^\tau)^T = \kappa^\tau \kappa A$$

and therefore $\kappa^\tau = \kappa^{-1}$. If we set $B = (1+\kappa)A$ then

$$(B^\tau)^T = (((1+\kappa)A)^\tau)^T = ((1+\kappa)^\tau)(A^\tau)^T = (1+\kappa^{-1})\kappa A = (\kappa+1)A = B.$$

The hyperplanes with coordinate vectors $(\mathbf{x}^\tau)^T A^T$ and $(\mathbf{x}^\tau)B^T$ are the same, for any vector $\mathbf{x}$. Hence, if $\kappa \neq -1$, we may take our polarity to be determined by a field automorphism τ with order dividing two and an invertible matrix B such that $(B^\tau)^T = B$. If $\kappa = -1$ then we observe that we may replace A by $C = \lambda A$ for any non-zero element of of $\mathbb{F}$. Then

$$(C^\tau)^T = -\frac{\lambda^\tau}{\lambda} C.$$

Thus if $\lambda^\tau/\lambda \neq 1$ we may replace A by C and then reapply our trick above to get a matrix B such that $(B^\tau)^T = B$. Problems remain only if $\lambda^\tau = \lambda$ for all elements λ of $\mathbb{F}$. But then τ must be the identity automorphism and $A^T = -A$. Our results can be summarised as follows.

7.3.1 Theorem. *Let ϕ be a polarity of $PG(n-1,\mathbb{F})$. Then there is an invertible $n \times n$ matrix A and a field automorphism τ such that $\mathbf{x}\phi = (\mathbf{x}^\tau)^T A$. Further, either*

(a) $(A^\tau)^T = A$ and τ has order two,

(b) $A^T = A$ and $\tau = 1$, or

(c) $A^T = -A$, the diagonal entries of A are zero and $\tau = 1$. □

The three types of polarity are known respectively as *Hermitian*, *orthogonal* and *symplectic*. The last two cases are not disjoint in characteristic two; a polarity that is both orthogonal and symplectic is usually be treated as symplectic. Our argument has actually established that polarities of these types exist—we need only choose an invertible matrix A and an optional field automorphism of order two.

7.4 Polar Spaces

Suppose that ϕ is a polarity of $PG(n,\mathbb{F})$. If H is a hyperplane then $p \in \cap\{u^\phi : u \in H\}$ if and only if $H = p^\phi$, or equivalently, if and only if $p = H^\phi$. If U is a subspace, we may therefore define

$$U^\phi = \bigcap_{u \in U} u^\phi.$$

A subspace U is *isotropic* if $U \subseteq U^\phi$. Any polarity thus determines a collection of isotropic spaces of $PG(n,\mathbb{F})$. (A point is isotropic if it is absolute.) The set of isotropic points of a polarity ϕ, together with the collection of its isotropic subspaces, provides the canonical example of a *polar space*. A polar space of rank n consists of a set of points S, together with a collection of subsets of S, called *subspaces*, such that the following axioms hold.

(a) A subspace, together with the subspaces it contains, forms a generalised projective space of rank at most n. (A generalised projective space is either a projective space, or consists of a set with all two-element subsets as lines.)

(b) Given a subspace U of rank n and a point p not in U, there is a unique subspace V which contains p and all points of U which are joined to p by a line; $\operatorname{rk}(U \cap V) = n - 1$.

(c) There are two disjoint subspaces of rank n.

A polar space is not a linear space, since there are pairs of points which are not collinear and the entire point set is not a subspace. However polar spaces make perfectly good matroids, as we will see. From Lemma 6.2.2 and Lemma 6.2.4, we see that every quadric is a polar space. In the next section we will study the connection between quadrics and polarities. To prove that the isotropic subspaces of an arbitrary polarity form a polar space requires some work. (The difficulty is to verify that there are pairs of disjoint maximal isotropic subspaces.)

There is an alternative, and simpler, approach to polar spaces. We define a *Shult space* to be an incidence structure $\mathcal{S}$ with the property that if p is a point and ℓ is line not on p then p is collinear with one, or all, the points on ℓ. A Shult space is not automatically a linear space, because we have not required that any two points lie on at most one line. A Shult space is *non-degenerate* if there is no point which is collinear to all the others. Buekenhout and Shult proved that any non-degenerate Shult space is a polar space. (The converse is an easy exercise.) We will present a proof of their result later. Note, however, that it is trivial to verify that the isotropic points and lines of a polarity form a Shult space, and hence a polar space. In the next few sections we consider the properties of the classical, finite, examples of polar spaces. Following this we will make an axiomatic study of polar spaces, including a proof of the Buekenhout-Shult theorem.

7.5 Quadratic Spaces and Polarities

Let V be a vector space over $\mathbb{F}$. A *quadratic form* Q over $\mathbb{F}$ is a function from V to $\mathbb{F}$ such that

(a) $Q(\lambda u) = \lambda^2 Q(u)$ for all λ in $\mathbb{F}$ and u in V, and

(b) $Q(u+v) - Q(u) - Q(v)$ is bilinear.

Let β be the bilinear form defined by

$$\beta(u,v) = Q(u+v) - Q(u) - Q(v).$$

We say that β is obtained from Q by *polarisation*. The above conditions imply that

$$4Q(u) = Q(2u) = Q(u+u) = 2Q(u) + \beta(u,u)$$

whence we have $\beta(u,u) = 2Q(u)$. Thus, if the characteristic of $\mathbb{F}$ is not even, the quadratic form is determined by β. If the characteristic of $\mathbb{F}$ is even then $\beta(u,u) = 0$ for all u in V. In this case we say that the form is *symplectic*. Each homogeneous quadratic polynomial in n variables over $\mathbb{F}$ determines a quadratic form on $\mathbb{F}^n$.

A quadratic form is *non-singular* if, when $Q(a) = 0$ and $\beta(a,v) = 0$ for all v, then $v = 0$. In odd characteristic a quadratic form is non-singular if and only if β is non-degenerate. (Exercise.) A subspace U of V is *singular* if $Q(u) = 0$ for all u in U.

We are going to classify quadratic forms over finite fields. One consequence of this will be the classification of orthogonal polarities over fields of odd characteristic. For any subspace W of V, we define

$$W^\perp = \{v \in V : \beta(v,w) = 0 \ \forall w \in W\}$$

If S is a subset of V we write $\langle S \rangle$ to denote the subspace spanned by V. If w is a vector in V then we will normally write $w^\perp$ rather than $\langle w \rangle^\perp$.

We define an *quadratic space* to be a pair (V,Q) where V is a vector space and Q is a quadratic form on V. We say that (V,Q) is non-singular if Q is. If (V,Q) is a quadratic space and U is a subspace of V, then (U,Q) is a quadratic space. This may be singular even if (V,Q) is not—for example, let U be the span of a singular vector. The form on (U,Q) is actually the restriction of Q to U, and should be denoted by $Q\restriction U$.

We note the following result, the proof of which is left as an exercise.

7.5.1 Lemma. *If W is a subspace of the quadratic space (V,Q), then the quotient space $W^\perp/W \cap W^\perp$ is an quadratic space with quadratic form $\overline{Q}$ satisfying $\overline{Q}(v+W) = Q(v)$.* □

Suppose (U, Q_U) and (V, Q_V) are quadratic spaces over $\mathbb{F}$. If $W := U \oplus V$, then the function Q_W defined by

$$Q_W((u,v)) = Q_U(u) + Q_V(v)$$

is a quadratic form on W. (It may be best to view this as follows: if $w \in W$ then we can express w uniquely as $w = u + v$ where $u \in U$ and $v \in V$, then $Q_W(w)$ is defined to be $Q_U(u) + Q_V(v)$.) The form Q_W is non-singular if and only if Q_U and Q_V are.

7.5.2 Lemma. *If W is a subspace of the quadratic space (V,Q) and $W \cap W^\perp = \{0\}$, then (V,Q) is the direct sum of the spaces (W,Q) and $(W^\perp, Q)$. If (V,Q) is non-singular, so are (W,Q) and $(W^\perp,Q)$.* □

An quadratic space is *anisotropic* if $Q(v) \neq 0$ for all non-zero vectors v in V. You may show that if a subspace (W,Q) of (V,Q) is anisotropic, then $W \cap W^\perp = \{0\}$.

7.5.3 Lemma. *If V is an anisotropic quadratic space over $GF(q)$ then $\dim V \leq 2$. If $\dim V = 2$ then V has a basis $\{d, d'\}$ such that $Q(d') = (d, d') = 1$.*

Proof. Assume that $\dim V \geq 2$. Choose a non-zero vector e in V and a vector d not in $e^\perp$. Let $W = \langle d, e\rangle$. Assume $Q(e) = \epsilon$ and that d has been chosen so that $(d,e) = \epsilon$. Assume further that $\sigma = Q(d)/\epsilon$. Then

$$Q(\alpha e + \beta d) = \alpha^2\epsilon + \beta^2\sigma\epsilon + \alpha\beta\epsilon = \epsilon(\alpha^2 + \alpha\beta + \beta^2\sigma).$$

If W is anisotropic then $\alpha^2 + \alpha\beta + \beta^2\sigma \neq 0$ for all α in $\mathbb{F}$. Hence the polynomial $x^2 + x + \sigma$ is irreducible over $\mathbb{F} = GF(q)$. Let θ be a root of it in $GF(q^2)$ and let $a \mapsto \bar{a}$ be the involutory automorphism of $\mathbb{F}(\theta)$. Then

$$Q(\alpha e + \beta d) = \epsilon(\alpha + \beta\theta)(\alpha + \beta\bar{\theta})$$

from which it follows that $\{Q(w) : w \in W\} = \mathbb{F}$. This means that we can assume that e was chosen so that $\epsilon = 1$. Finally, if $n \geq 3$ and v is a non-zero vector in $\langle d, e\rangle^\perp$ then $Q(v) = -Q(w)$ for some w in V. Then $Q(v + w) = 0$ and V is not anisotropic. □

It follows readily from the above lemma that, up to isomorphism, there is only one anisotropic quadratic space of dimension two over a finite field $\mathbb{F}$. We note, if $\mathbb{F}$ is finite and $Q(x) = 0$ for some x then $(x,x) = 0$. For if q is even then $(x,x) = 0$ for all x, and if q is odd then $0 = 2Q(x) = (x,x)$ again implies that $(x,x) = 0$.

7.5.4 Theorem. *Let (V,Q) be an quadratic space of dimension n over $GF(q)$. Then V has a basis of one the following forms:*

(a) $n = 2m :\ e_1, \dots, e_m; f_1, \dots, f_m$ *where*

$$Q(e_i) = Q(f_i) = 0,\ (e_i, f_j) = \delta_{ij},\ (e_i, e_j) = (f_i, f_j) = 0$$

(b) $n = 2m+2: \ d, d', e_1, \ldots, e_m; f_1, \ldots, f_m$ *with the* e_i *and* f_j *as in (a),* $\langle d, d' \rangle$ *an anisotropic quadratic space with* $Q(d') = (d, d') = 1$, $Q(d) = \sigma$ *where* $x^2 + x + \sigma$ *is irreducible over* $GF(q)$ *and*

$$(d, e_i) = d(f_i) = (d', e_i) = (d', f_i) = 0$$

(c) $n = 2m+1 : d, e_1, \ldots, e_m; f_1, \ldots, f_m$ *and everything as in (b).*

Proof. Assume that $\dim V \geq 3$, and let e_1 be a non-zero vector in V with $Q(e_1) = 0$. Then there is a vector f in V such that $(e_1, f) = 1$ and

$$Q(\alpha e_1 + f) = Q(f) + \alpha.$$

If we set f_1 equal to $-Q(f)e_1 + f$ then $Q(f_1) = 0$ and $(e_1, f_1) = 1$. (Here we are using the fact that $(e_1, e_1) = 0$.) Now V is the orthogonal direct sum of $\langle e_1, f_1 \rangle$ and $\langle e_1, f_1 \rangle^{\perp}$, and the result follows by induction. □

We can write down the quadratic forms corresponding to the three cases of the theorem as follows:

(a) $Q(\sum \alpha_i e_i + \sum \beta_i f_i) = \sum \alpha_i \beta_i$

(b) $Q(\gamma d + \gamma' d' + \sum \alpha_i e_i + \sum \beta_i f_i) = \gamma^2 \sigma + \gamma\gamma' + {\gamma'}^2 + \sum \alpha_i \beta_i$

(c) $Q(\gamma d + \sum \alpha_i e_i + \sum \beta_i f_i) = \gamma^2 \sigma + \sum \alpha_i \beta_i$

In both (b) and (c), the field element σ is chosen so that $x^2 + x + \sigma$ is irreducible over $GF(q)$.

An isometry of the quadratic space (V, Q) is an element τ of $GL(V)$ such that $Q(v\tau) = Q(v)$ for all v in V. The set of all isometries of V is the *isometry group* of V. It is denoted by $O(V)$ in general, and by $O^+(2m, q)$, $O^-(2m+2, q)$ and $O(2m+1, q)$ respectively in cases (a), (b) and (c) above.