Projective and Polar Spaces

Peter J. Cameron
Queen Mary and Westfield College
2000

Preface

It is common now in academic circles to lament the decline in the teaching of geometry in our schools and universities, and the resulting loss of "geometric intuition" among our students. On the other hand, recent decades have seen renewed links between geometry and physics, to the benefit of both disciplines. One of the world's leading mathematicians has argued that the insights of "pre-calculus" geometry have a rôle to play at all levels of mathematical activity (Arnol'd [A]). There is no doubt that a combination of the axiomatic and the descriptive approaches associated with algebra and geometry respectively can help avoid the worst excesses of either approach alone.

These notes are about geometry, but by no means all or even most of geometry. I am concerned with the geometry of incidence of points and lines, over an arbitrary field, and unencumbered by metrics or continuity (or even betweenness). The major themes are the projective and affine spaces, and the polar spaces associated with sesquilinear or quadratic forms on projective spaces. The treatment of these themes blends the descriptive (*What do these spaces look like?*) with the axiomatic (*How do I recognize them?*) My intention is to explain and describe, rather than to give detailed argument for every claim. Some of the theorems (especially the characterisation theorems) are long and intricate. In such cases, I give a proof in a special case (often over the field with two elements), and an outline of the general argument.

The classical works on the subject are the books of Dieudonné [L] and Artin [B]. I do not intend to compete with these books. But much has happened since they were written (the axiomatisation of polar spaces by Veldkamp and Tits (see Tits [S]), the classification of the finite simple groups with its many geometric spin-offs, Buekenhout's geometries associated with diagrams, etc.), and I have included some material not found in the classical books.

Roughly speaking, the first five chapters are on projective spaces, the last five on polar spaces. In more detail: Chapter 1 introduces projective and affine

spaces synthetically, and derives some of their properties. Chapter 2, on projective planes, discusses the rôle of Desargues' and Pappus' theorems in the coordinatisation of planes, and gives examples of non-Desarguesian planes. In Chapter 3, we turn to the coordinatisation of higher-dimensional projective spaces, following Veblen and Young. Chapter 4 contains miscellaneous topics: recognition of some subsets of projective spaces, including conics over finite fields of odd characteristic (Segre's theorem); the structure of projective lines; and generation and simplicity of the projective special linear groups. Chapter 5 outlines Buekenhout's approach to geometry via diagrams, and illustrates by interpreting the earlier characterisation theorems in terms of diagrams.

Chapter 6 relates polarities of projective spaces to reflexive sesquilinear forms, and gives the classification of these forms. Chapter 7 defines polar spaces, the geometries associated with such forms, and gives a number of these properties; the Veldkamp–Tits axiomatisation and the variant due to Buekenhout and Shult are also discussed, and proved for hyperbolic quadrics and for quadrics over the 2-element field. Chapter 8 discusses two important low-dimensional phenomena, the Klein quadric and triality, proceeding as far as to define the polarity defining the Suzuki–Tits ovoids and the generalised hexagon of type G_2. In Chapter 9, we take a detour to look at the geometry of the Mathieu groups. This illustrates that there are geometric objects satisfying axioms very similar to those for projective and affine spaces, and also having a high degree of symmetry. In the final chapter, we define spinors and use them to investigate the geometry of dual polar spaces, especially those of hyperbolic quadrics.

The notes are based on postgraduate lectures given at Queen Mary and Westfield College in 1988 and 1991. I am grateful to members of the audience on these occasions for their comments and especially for their questions, which forced me to think things through more carefully than I might have done. Among many pleasures of preparing these notes, I count two lectures by Jonathan Hall on his beautiful proof of the characterisation of quadrics over the 2-element field, and the challenge of producing the diagrams given the constraints of the typesetting system!

In the introductory chapters to both types of spaces (Chapters 1 and 6), as well as elsewhere in the text (especially Chapter 10), some linear algebra is assumed. Often, it is necessary to do linear algebra over a non-commutative field; but the differences from the commutative case are discussed. A good algebra textbook (for example, Cohn (1974)) will contain what is necessary.

Peter J. Cameron, London, 1991

Preface to the second edition

Materially, this edition is not very different from the first edition which was published in the QMW Maths Notes series in 1991. I have converted the files into LaTeX, corrected some errors, and added some new material and a few more references; this version does not represent a complete bringing up-to-date of the original. I intend to publish these notes on the Web.

In the meantime, one important relevant reference has appeared: Don Taylor's book *The Geometry of the Classical Groups* [R]. (Unfortunately, it has already gone out of print!) You can also look at my own lecture notes on Classical Groups (which can be read in conjunction with these notes, and which might be integrated with them one day). Other sources of information include the *Handbook of Incidence Geometry* [E] and (on the Web) two series of SOCRATES lecture notes at
`http://dwispc8.vub.ac.be/Potenza/lectnotes.html`
and
`http://cage.rug.ac.be/~fdc/intensivecourse2/final.html`

Please note that, in Figure 2.3, there are a few lines missing: dotted lines *utq* and *urv* and a solid line ub_1c_2. (The reason for this is hinted at in Exercise 3 in Section 1.2.)

Peter J. Cameron, London, 2000

Contents

1

Projective spaces

In this chapter, we describe projective and affine spaces synthetically, in terms of vector spaces, and derive some of their geometric properties.

1.1 Fields and vector spaces

Fields will not necessarily be commutative; in other words, the term "field" will mean "division ring" or "skew field", while the word "commutative" will be used where necessary. Often, though, I will say "skew field", as a reminder. (Of course, this refers to the multiplication only; addition will always be commutative.)

Given a field F, let

$$I = \{n \in \mathbb{N} : (\forall \alpha \in F)\ n \cdot \alpha = 0\} = \{n \in \mathbb{N} : n \cdot 1_F = 0\}.$$

Then I is an ideal in $\mathbb{N}$, hence $I = (c)$ for some non-negative integer c called the *characteristic* of F. The characteristic is either 0 or a prime number. For each value of the characteristic, there is a unique *prime field* which is a subfield of any field of that characteristic: the rational numbers in characteristic zero, and the integers modulo p in prime characteristic p.

Occasionally I will assume rudimentary results about field extensions, degree, and so on.

Much of the time, we will be concerned with finite fields. The main results about these are as follows.

Theorem 1.1 (Wedderburn's Theorem) *A finite field is commutative.*

Theorem 1.2 (Galois' Theorem) *A finite field has prime power order. For any prime power q, there is a unique finite field of order q.*

The unique field of order q is denoted by $\mathrm{GF}(q)$. If $q = p^d$ with p prime, its additive structure is that of a d-dimensional vector space over its prime field $\mathrm{GF}(p)$ (the integers modulo p). Its multiplicative group is cyclic (of order $q-1$), and its automorphism group is cyclic (of order d). If $d = 1$ (that is, if q is prime), then $\mathrm{GF}(q)$ is the ring of integers mod q.

An *anti-automorphism* of a field is a bijection σ with the properties

$$\begin{aligned}(c_1+c_2)^\sigma &= c_1^\sigma + c_2^\sigma,\\ (c_1\cdot c_2)^\sigma &= c_2^\sigma \cdot c_1^\sigma.\end{aligned}$$

The identity (or, indeed, any automorphism) is an anti-automorphism of a commutative field. Some non-commutative fields have anti-automorphisms. A well-known example is the field $\mathbb{H}$ of quaternions, with a basis over $\mathbb{R}$ consisting of elements $1, i, j, k$ satisfying

$$i^2 = j^2 = k^2 = -1, \qquad ij = k, \quad jk = i, \quad ki = j;$$

the anti-automorphism is given by

$$a + bi + cj + dk \mapsto a - bi - cj - dk.$$

Others, however, do not.

The *opposite* of the field $(F, +, \cdot)$ is the field $(F, +, \circ)$, where the binary operation $\circ$ is defined by the rule

$$c_1 \circ c_2 = c_2 \cdot c_1.$$

Thus, an anti-automorphism of F is just an isomorphism between F and its opposite F°.

For non-commutative fields, we have to distinguish between left and right vector spaces. In a left vector space, if we write the product of the scalar c and the vector $\mathbf{v}$ as $c\mathbf{v}$, then $c_1(c_2\mathbf{v}) = (c_1c_2)\mathbf{v}$ holds. In a right vector space, this condition reads $c_1(c_2\mathbf{v}) = (c_2c_1)\mathbf{v}$. It is more natural to write the scalars on the right (thus: $\mathbf{v}c$), so that the condition is $(\mathbf{v}c_2)c_1 = \mathbf{v}(c_2c_1)$). A right vector space over F is a left vector space over F°.

Our vector spaces will almost always be finite dimensional.

For the most part, we will use left vector spaces. In this case, it is natural to represent a vector by the row tuple of its coordinates with respect to some basis; scalar multiplication is a special case of matrix multiplication. If the vector space has dimension n, then vector space endomorphisms are represented by $n \times n$ matrices, acting on the right, in the usual way:

$$(\mathbf{v}A) = \sum_i v_i A_{ij}$$

if $\mathbf{v} = (v_1, \ldots, v_n)$.

The *dual space* V^* of a (left) vector space V is the set of linear maps from V to F, with pointwise addition and with scalar multiplication defined by

$$(\mathbf{f}c)\mathbf{v} = \mathbf{f}(c\mathbf{v}).$$

Note that this definition makes V^* a right vector space.

1.2 Projective spaces

A projective space of dimension n over a field F (not necessarily commutative!) can be constructed in either of two ways: by adding a hyperplane at infinity to an affine space, or by "projection" of an $(n+1)$-dimensional space. Both methods have their importance, but the second is the more natural.

Thus, let V be an $(n+1)$-dimensional left vector space over F. The *projective space* $\mathrm{PG}(n,F)$ is the geometry whose *points, lines, planes,* ... are the vector subspaces of V of dimensions 1, 2, 3, ...

Note that the word "geometry" is not defined here; the properties which are regarded as geometrical will emerge during the discussion.

Note also the dimension shift: a d-dimensional projective subspace (or flat) is a $(d+1)$-dimensional vector subspace. This is done in order to ensure that familiar geometrical properties hold. For example, two points lie on a unique line; two intersecting lines lie in a unique plane; and so on. Moreover, any d-dimensional projective subspace is a d-dimensional projective space in its own right (when equipped with the subspaces it contains).

To avoid confusion (if possible), I will from now on reserve the term *rank* (in symbols, rk) for vector space dimension, so that unqualified "dimension" will be geometric dimension.

A *hyperplane* is a subspace of codimension 1 (that is, of dimension one less than the whole space). If H is a hyperplane and L a line not contained in H, then $H \cap L$ is a point.

A projective plane (that is, $\mathrm{PG}(2,F)$) has the property that any two lines meet in a (unique) point. For, if $\mathrm{rk}(V) = 3$ and $U, W \subseteq V$ with $\mathrm{rk}(U) = \mathrm{rk}(W) = 2$, then $U + W = V$, and so $\mathrm{rk}(U \cap W) = 1$; that is, $U \cap W$ is a point. From this, we deduce:

Proposition 1.3 (Veblen's Axiom) *If a line intersects two sides of a triangle but doesn't contain their intersection, then it intersects the third side also.*

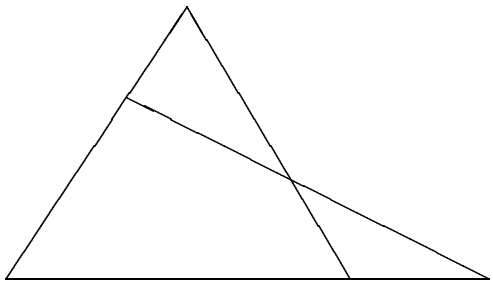


Figure 1.1: Veblen's Axiom

For the triangle is contained in a plane, and the hypotheses guarantee that the line in question is spanned by points in the plane, and hence also lies in the plane.

Veblen's axiom is sometimes called the Veblen-Young Axiom or Pasch's Axiom. The latter name is not strictly accurate: Pasch was concerned with real projective space, and the fact that if two intersections are inside the triangle, the third is outside; this is a property involving order, going beyond the incidence geometry which is our concern here. In Section 3.1 we will see why 1.3 is referred to as an "axiom".

Another general geometric property of projective spaces is the following.

Proposition 1.4 (Desargues' Theorem) *In Figure 1.2, the three points p, q, r are collinear.*

In the case where the figure is not contained in a plane, the result is obvious geometrically. For each of the three points p, q, r lies in both the planes $a_1b_1c_1$ and $a_2b_2c_2$; these planes are distinct, and both lie in the 3-dimensional space spanned by the three lines through o, and so their intersection is a line.

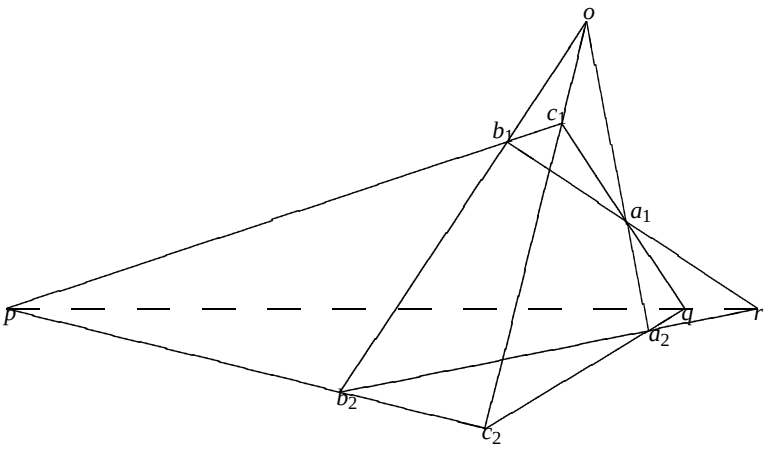


Figure 1.2: Desargues' Theorem

The case where the figure is contained in a plane can be deduced from the "general" case as follows. Given a point o and a hyperplane H,write $aa' \sim bb'$ if oaa', obb' are collinear triples and the lines ab and $a'b'$ intersect in H (but none of the points a, a', b, b' lies in H). Now Desargues' Theorem is the assertion that the relation $\sim$ is transitive. (For p, q, r are collinear if and only if every hyperplane containing p and q also contains r; it is enough to assume this for the hyperplanes not containing the points a, a', etc.) So suppose that $aa' \sim bb' \sim cc'$. The geometric argument of the preceding paragraph shows that $aa' \sim cc'$ if the configuration is not coplanar; so suppose it is. Let od be a line not in this plane, with $d \notin H$, and choose d' such that $aa' \sim dd'$. Then $bb' \sim dd'$, $cc' \sim dd'$, and $aa' \sim cc'$ follow in turn from the non-planar Desargues' Theorem.

(If we are only given a plane initially, the crucial fact is that the plane can be embedded in a 3-dimensional space.)

Remark The case where $|F| = 2$ is not covered by this argument — can you see why? — and, indeed, the projective plane over $\mathrm{GF}(2)$ contains no non-degenerate Desargues configuration: it only contains seven points! Nevertheless, Desargues' Theorem holds, in the sense that any meaningful degeneration of it is true in the projective plane over $\mathrm{GF}(2)$. We will not make an exception of this case.

It is also possible to prove Desargues' Theorem algebraically, by choosing

coordinates (see Exercise 1). However, it is important for later developments to know that a purely geometric proof is possible.

Let V be a vector space of rank $n+1$ over F, and V^* its dual space. As we saw, V^* is a right vector space over F, and so can be regarded as a left vector space over the opposite field F°. It has the same rank as V if this is finite. Thus we have projective spaces $\mathrm{PG}(n,F)$ and $\mathrm{PG}(n,F^\circ)$, standing in a dual relation to one another. More precisely, we have a bijection between the flats of $\mathrm{PG}(n,F)$ and those of $\mathrm{PG}(n,F^\circ)$, given by

$$U \leftrightarrow \mathrm{Ann}(U) = \{\mathbf{f} \in V^* : (\forall \mathbf{u} \in U)\ (\mathbf{fu} = 0)\}.$$

This correspondence preserves incidence and reverses inclusion:

$$\begin{aligned} U_1 \subseteq U_2 &\Rightarrow \mathrm{Ann}(U_2) \subseteq \mathrm{Ann}(U_1), \\ \mathrm{Ann}(U_1 + U_2) &= \mathrm{Ann}(U_1) \cap \mathrm{Ann}(U_2), \\ \mathrm{Ann}(U_1 \cap U_2) &= \mathrm{Ann}(U_1) + \mathrm{Ann}(U_2). \end{aligned}$$

Moreover, the (geometric) dimension of $\mathrm{Ann}(U)$ is $n - 1 - \dim(U)$.

This gives rise to a *duality principle*, where any configuration theorem in projective space translates into another (over the opposite field) in which inclusions are reversed and dimensions suitably modified. For example, in the plane, the dual of the statement that two points lie on a unique line is the statement that two lines meet in a unique point.

We turn briefly to affine spaces. The description closest to that of projective spaces runs as follows. Let V be a vector space of rank n over F. The *points, lines, planes, ...* of the *affine space* $\mathrm{AG}(n,F)$ are the cosets of the vector subspaces of rank 0, 1, 2, ... (No dimension shift this time!) In particular, points are cosets of the zero subspace, in other words, singletons, and we can identify them with the vectors of V. So the affine space is "a vector space with no distinguished origin".

The other description is: $\mathrm{AG}(n,F)$ is obtained from $\mathrm{PG}(n,F)$ by deleting a hyperplane together with all the subspaces it contains.

The two descriptions are matched up as follows. Take the vector space

$$V = F^{n+1} = \{(x_0, x_1, \ldots, x_n) : x_0, \ldots, x_n \in F\}.$$

Let W be the hyperplane defined by the equation $x_0 = 0$. The points remaining are rank 1 subspaces spanned by vectors with $x_0 \neq 0$; each point has a unique spanning vector with $x_0 = 1$. Then the correspondence between points in the two descriptions is given by

$$\langle (1, x_1, \ldots, x_n) \rangle \leftrightarrow (x_1, \ldots, x_n).$$

(See Exercise 2.)

In $\mathrm{AG}(n,F)$, we say that two subspaces are *parallel* if (in the first description) they are cosets of the same vector subspace, or (in the second description) they have the same intersection with the deleted hyperplane. Parallelism is an equivalence relation. Now the projective space can be recovered from the affine space as follows. To each parallel class of d-dimensional subspaces of $\mathrm{AG}(n,F)$ corresponds a unique $(d-1)$-dimensional subspace of $\mathrm{PG}(n-1,F)$. Adjoin to the affine space the points (and subspaces) of $\mathrm{PG}(n-1,F)$, and adjoin to all members of a parallel class all the points in the corresponding subspace. The result is $\mathrm{PG}(n,F)$.

The distinguished hyperplane is called the *hyperplane at infinity* or *ideal hyperplane*. Thus, an affine space can also be regarded as "a projective space with a distinguished hyperplane".

The study of projective geometry is in a sense the outgrowth of the Renaissance theory of perspective. If a painter, with his eye at the origin of Euclidean 3-space, wishes to represent what he sees on a picture plane, then each line through the origin (i.e., each rank 1 subspace) should be represented by a point of the picture plane, viz., the point at which it intersects the picture plane. Of course, lines parallel to the picture plane do not intersect it, and must be regarded as meeting it in ideal "points at infinity". Thus, the physical picture plane is an affine plane, and is extended to a projective plane; and the points of the projective plane are in one-to-one correspondence with the rank 1 subspaces of Euclidean 3-space. It is easily checked that lines of the picture plane correspond to rank 2 subspaces, provided we make the convention that the points at infinity comprise a single line. Not that the picture plane really is affine rather than Euclidean; the ordinary distances in it do not correspond to distances in the real world.

Exercises

1. Prove Desargues' Theorem in coordinates.

2. Show that the correspondence defined in the text between the two descriptions of affine space is a bijection which preserves incidence, dimension, and parallelism.

3. The LaTeX typesetting system provides facilities for drawing diagrams. In a diagram, the slope of a line is restricted to being infinity or a rational number whose numerator and denominator are each at most 6 in absolute value.

(a) What is the relation between the slopes of the six lines of a complete quadrangle (all lines joining four points)? Investigate how such a figure can be

drawn with the above restriction on the slopes.

(b) Investigate similarly how to draw a Desargues configuration.

1.3 The "Fundamental Theorem of Projective Geometry"

An *isomorphism* between two projective spaces is a bijection between the point sets of the spaces which maps any subspace into a subspace (when applied in either direction). A *collineation* of $\mathrm{PG}(n,F)$ is an isomorphism from $\mathrm{PG}(n,F)$ to itself. The theorem of the title of this section has two consequences: first, that isomorphic projective spaces have the same dimension and the same coordinatising field; second, a determination of the group of all collineations.

We must assume that $n > 1$; for the only proper subspaces of a projective line are its points, and so any bijection is an isomorphism, and the collineation group is the full symmetric group. (There are methods for assigning additional structure to a projective line, for example, using cross-ratio; these will be discussed later on, in Section 4.5.)

The *general linear group* $\mathrm{GL}(n+1,F)$ is the group of all non-singular linear transformations of $V = F^{n+1}$; it is isomorphic to the group of invertible $(n+1) \times (n+1)$ matrices over F. (In general, the determinant is not well-defined, so we cannot identify the invertible matrices with those having non-zero determinant.) Any element of $\mathrm{GL}(n+1,F)$ maps subspaces of V into subspaces of the same rank, and preserves inclusion; so it induces a collineation of $\mathrm{PG}(n,F)$. The group $\mathrm{Aut}(F)$ of automorphisms of F has a coordinate-wise action on V^{n+1}; these transformations also induce collineations. The group generated by $\mathrm{GL}(n+1,F)$ and $\mathrm{Aut}(F)$ (which is actually their semi-direct product) is denoted by $\Gamma\mathrm{L}(n+1,F)$; its elements are called *semilinear transformations*. The groups of collineations of $\mathrm{PG}(n,F)$ induced by $\mathrm{GL}(n+1,F)$ and $\Gamma\mathrm{L}(n+1,F)$ are denoted by $\mathrm{PGL}(n+1,F)$ and $\mathrm{P\Gamma L}(n+1,F)$, respectively.

More generally, a semi-linear transformation from one vector space to another is the composition of a linear transformation and a coordinate-wise field automorphism of the target space.

Theorem 1.5 (Fundamental Theorem of Projective Geometry) *Any isomorphism between projective spaces of dimension at least* 2 *is induced by a semilinear transformation between the underlying vector spaces, unique up to scalar multiplication.*

Before outlining the proof, we will see the two important corollaries of this result. Both follow immediately from the theorem (in the second case, by taking the two projective spaces to be the same).

Corollary 1.6 *Isomorphic projective spaces of dimension at least* 2 *have the same dimension and are coordinatised by isomorphic fields.* ■

Corollary 1.7 *(a) For* $n > 1$*, the collineation group of* $\mathrm{PG}(n,F)$ *is the group* $\mathrm{P\Gamma L}(n+1,F)$.

(b) The kernel of the action of $\Gamma\mathrm{L}(n+1,F)$ *on* $\mathrm{PG}(n,F)$ *is the group of non-zero scalars (acting by left multiplication).* ■

Remark The point of the theorem, and the reason for its name, is that the algebraic structure of the underlying vector space can be recovered from the incidence geometry of the projective space. The proof is a good warm-up for the coordinatisation theorems I will be discussing soon. In fact, the proof concentrates on Corollary 1.7, for ease of exposition. The dimension of a projective space is two less than the number of subspaces in a maximal chain (under inclusion); and our argument shows that the geometry determines the coordinatising field up to isomorphism.

Proof We show first that two semi-linear transformations which induce the same collineation differ only by a scalar factor. By following one by the inverse of the other, we see that it suffices to show that a semi-linear transformation which fixes every point of $\mathrm{PG}(n,F)$ is a scalar multiplication. So let $\mathbf{v} \mapsto \mathbf{v}^\sigma A$ fix every point of $\mathrm{PG}(n,F)$, where $\sigma \in \mathrm{Aut}(F)$ and $A \in \mathrm{GL}(n+1,F)$. Then every vector is mapped to a scalar multiple of itself. Let $\mathbf{e}_0,\ldots,\mathbf{e}_n$ be the standard basis for V. Then (since σ fixes the standard basis vectors) we have $\mathbf{e}_i A = \lambda_i \mathbf{e}_i$ for $i = 0,\ldots,n$. Also,

$$\begin{aligned}(\mathbf{e}_0 + \ldots + \mathbf{e}_n)A &= \lambda_0\mathbf{e}_0 + \ldots + \lambda_n\mathbf{e}_n \\ &= \lambda(\mathbf{e}_0 + \ldots + \mathbf{e}_n), \quad \text{say,}\end{aligned}$$

so $\lambda_0 = \ldots = \lambda_n = \lambda$.

Now, for any $\mu \in F$, the vector $(1,\mu,0,\ldots,0)$ is mapped to the vector $(\lambda,\mu^\sigma\lambda,0,\ldots,0)$; so we have $\lambda\mu = \mu^\sigma\lambda$. Thus

$$\mathbf{v}^\sigma A = \mathbf{v}^\sigma\lambda = \lambda\mathbf{v}$$

for any vector $\mathbf{v}$, as required.

Note that the field automorphism σ is conjugation by the element λ (that is, $\mu^\sigma = \lambda\mu\lambda^{-1}$); in other words, an inner automorphism.

Now we prove that any isomorphism is semilinear. The strategy is similar. Call an $(n+2)$ tuple of points *special* if no $n+1$ of them are linearly dependent. We have:

> There is a linear map carrying any special tuple to any other (in the same space, or another space of the same dimension over the same field).

(For, given a special tuple in the first space, spanning vectors for the first $n+1$ points form a basis $\mathbf{e}_0,\ldots,\mathbf{e}_n$, and the last point is spanned by a vector with all coordinates non-zero relative to this basis. Adjusting the basis vectors by scalar factors, we may assume that the last point is spanned by $\mathbf{e}_0+\ldots+\mathbf{e}_n$. Similarly, the points of a special tuple in the second space are spanned by the vectors of a basis $\mathbf{f}_0,\ldots,\mathbf{f}_n$, and $\mathbf{f}_0+\ldots+\mathbf{f}_n$. The unique linear transformation carrying the first basis to the second also carries the first special tuple to the second.)

Let θ be any isomorphism. Then there is a linear map ϕ which mimics the effect of θ on a special $(n+2)$-tuple. Composing θ with the inverse of ϕ, we obtain an automorphism of $\mathrm{PG}(n,F)$ which fixes the $(n+2)$-tuple pointwise. We have to show that such an automorphism is the product of a scalar and a field automorphism. (Note that, as we saw above, left and right multiplications by λ differ by an inner automorphism.)

We assume that $n=2$; this simplifies the argument, while retaining its essential features. So let g be a collineation fixing the spans of $\mathbf{e}_0,\mathbf{e}_1,\mathbf{e}_2$ and $\mathbf{e}_0+\mathbf{e}_1+\mathbf{e}_2$. We use homogeneous coordinates, writing these vectors as $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, and $(1,1,1)$, and denote the general point by (x,y,z).

The points on the line $\{(x_0,0,x_2)\}$, apart from $(1,0,0)$, have the form $(x,0,1)$ for $x\in F$, and so can be identified with elements of F. Now the bijection between this set and the set of points $(0,y,1)$ on the line $\{(0,x_1,x_2)\}$, given by $(x,0,1)\mapsto(0,x,1)$, can be geometrically defined in a way which is invariant under collineations fixing the four reference points (see Fig. 1.3). The figure also shows that the coordinates of all points in the plane are determined.

Furthermore, the operations of addition and multiplication in F can be defined geometrically in the same sense (see Figures 1.4 and 1.6). (The definitions look more familiar if we take the line $\{(x_1,x_2,0)\}$ to be at infinity, and draw the figure in the affine plane with lines through $(1,0,0)$ and $(0,1,0)$ horizontal and vertical respectively. this has been done for addition in Figure 1.5; the reader should draw

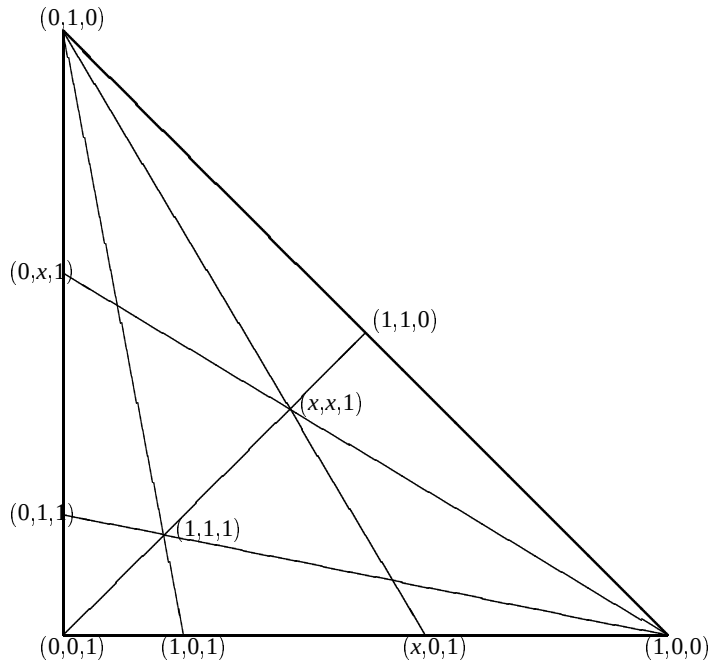


Figure 1.3: Bijection between the axes

the corresponding diagram for multiplication.) It follows that any collineation fixing our four basic points induces an automorphism of the field F, and its actions on the coordinates agree. The theorem is proved. ■

A group G acting on a set Ω is said to be *t-transitive* if, given any two t-tuples $(\alpha_1,\ldots,\alpha_t)$ and $(\beta_1,\ldots,\beta_t)$ of distinct elements of Ω, some element of G carries the first tuple to the second. G is *sharply t-transitive* if there is a unique such element. (If the action is not faithful, it is better to say: two elements of G which agree on t distinct points of Ω agree everywhere.)

Since any two distinct points of $\mathrm{PG}(n,F)$ are linearly independent, we see that $\mathrm{P\Gamma L}(n+1,F)$ (or even $\mathrm{PGL}(n+1,F)$) is 2-transitive on the points of $\mathrm{PG}(n,F)$. It is never 3-transitive (for $n>1$); for some triples of points are collinear and others are not, and no collineation can map one type to the other.

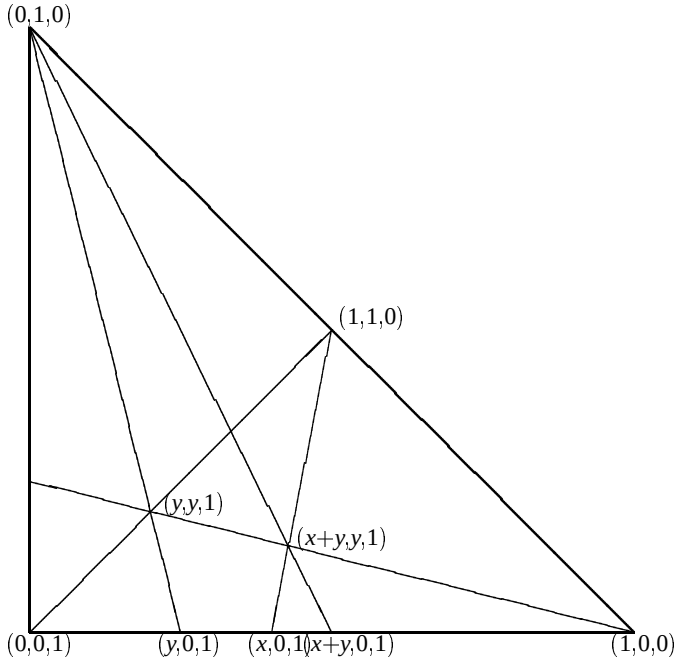


Figure 1.4: Addition

I will digress here to describe the analogous situation for $\mathrm{PG}(1,F)$, even though the FTPG does not apply in this case.

Proposition 1.8 *(a) The group* $\mathrm{PGL}(2,F)$ *is* 3*-transitive on the points of* $\mathrm{PG}(1,F)$*, and is sharply* 3*-transitive if and only if* F *is commutative.*

(b) There exist skew fields F *for which the group* $\mathrm{PGL}(2,F)$ *is* 4*-transitive on* $\mathrm{PG}(1,F)$*.*

Proof The first part follows just as in the proof of the FTPG, since any three points of $\mathrm{PG}(1,F)$ have the property that no two are linearly dependent. Again, as in that theorem, the stabiliser of the three points with coordinates $(1,0)$, $(0,1)$ and $(1,1)$ is the group of inner automorphisms of F, and so is trivial if and only if F is commutative.

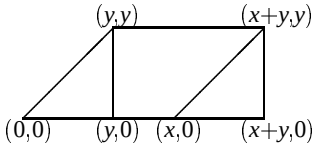


Figure 1.5: Affine addition

There exist skew fields F with the property that any two elements different from 0 and 1 are conjugate in the multiplicative group of F. Clearly these have the required property. (This fact is due to P. M. Cohn [15]; it is established by a construction analogous to that of Higman, Neumann and Neumann [20] for groups. Higman *et al.* used their construction to show that there exist groups in which all non-identity elements are conjugate; Cohn's work shows that there are multiplicative groups of skew fields with this property. Note that such a field has characteristic 2. For, if not, then $1+1 \neq 0$, and any automorphism must fix $1+1$.) ■

Finally, we consider collineations of affine spaces.

Parallelism in an affine space has an intrinsic, geometric definition. For two d-flats are parallel if and only if they are disjoint and some $(d+1)$-flat contains both. It follows that any collineation of $\mathrm{AG}(n,F)$ preserves parallelism. The hyperplane at infinity can be constructed from the parallel classes (as we saw in Section 1.2); so any collineation of $\mathrm{AG}(n,F)$ induces a collineation of this hyperplane, and hence of the embedding $\mathrm{PG}(n,F)$. Hence:

Theorem 1.9 *The collineation group of* $\mathrm{AG}(n,F)$ *is the stabiliser of a hyperplane in the collineation group of* $\mathrm{PG}(n,F)$. ■

Using this, it is possible to determine the structure of this group for $n > 1$ (see Exercise 2).

Proposition 1.10 *For* $n > 1$, *the collineation group of* $\mathrm{AG}(n,F)$ *is the semi-direct product of the additive group of* F^n *and* $\Gamma\mathrm{L}(n,F)$.

This group is denoted by $\mathrm{A}\Gamma\mathrm{L}(n,F)$. The additive group acts by translation, and the semilinear group in the natural way.

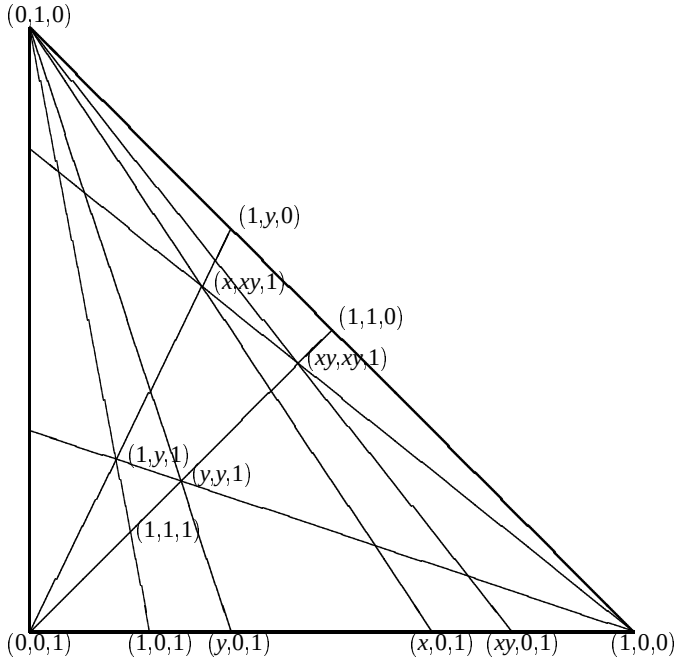


Figure 1.6: Multiplication

Exercises

1. Prove the FTPG for $n > 2$.

2. Use the correspondence between the two definitions of $\mathrm{AG}(n,F)$ given in the last section to deduce Proposition 1.10 from Theorem 1.9.

1.4 Finite projective spaces

Over the finite field $\mathrm{GF}(q)$, the n-dimensional projective and affine spaces and their collineation groups are finite, and can be counted. In this section, we display some of the relevant formulæ. We abbreviate $\mathrm{PG}(n,\mathrm{GF}(q))$ to $\mathrm{PG}(n,q)$, and similarly for affine spaces, collineation groups, etc.

A vector space of rank n over $\mathrm{GF}(q)$ is isomorphic to $\mathrm{GF}(q)^n$, and so the number of vectors is q^n. In consequence, the number of vectors outside a subspace

of rank k is $q^n - q^k$.

Proposition 1.11 *The number of subspaces of rank k in a vector space of rank n over $\mathrm{GF}(q)$ is*

$$\frac{(q^n-1)(q^n-q)\cdots(q^n-q^{k-1})}{(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})}.$$

Remark This number is called a *Gaussian coefficient*, and is denoted by $\begin{bmatrix} n \\ k \end{bmatrix}_q$.

Proof First we count the number of choices of k linearly independent vectors. The i^{th} vector may be chosen arbitrarily outside the subspace of rank $i-1$ spanned by its predecessors, hence in $q^n - q^{i-1}$ ways. Thus, the numerator is the required number of choices.

Now any k linearly independent vectors span a unique subspace of rank k; so the number of subspaces is found by dividing the number just calculated by the number of choices of a basis for a space of rank k. But the latter is given by the same formula, with k replacing n. ■

Proposition 1.12 *The order of $\mathrm{GL}(n,q)$ is*

$$(q^n-1)(q^n-q)\cdots(q^n-q^{n-1}).$$

The order of $\Gamma\mathrm{L}(n,q)$ is the above number multiplied by d, where $q = p^d$ with p prime; and the orders of $\mathrm{PGL}(n,q)$ and $\mathrm{P\Gamma L}(n,q)$ are obtained by dividing these numbers by $(q-1)$.

Proof An element of $\mathrm{GL}(n,q)$ is uniquely determined by the image of the standard basis, which is an arbitrary basis of $\mathrm{GF}(q)^n$; and the proof of Proposition 1.11 shows that the number of bases is the number quoted. The remainder of the proposition follows from the remarks in Section 1.3, since $\mathrm{GF}(q)$ has $q-1$ non-zero scalars, and its automorphism group has order d. ■

The formula for the Gaussian coefficient makes sense, not just for prime power values of q, but for any value of q different from 1. There is a combinatorial interpretation for any integer $q > 1$ (Exercise 3). Moreover, by l'Hôpital's rule, $\lim_{q\to 1}(q^a-1)/(q^b-1) = a/b$; it follows that

$$\lim_{q\to 1} \begin{bmatrix} n \\ k \end{bmatrix}_q = \binom{n}{k}.$$

This illustrates just one of the many ways in which subspaces of finite vector spaces resemble subsets of sets.

It follows immediately from Propopsition 1.11 that the numbers of k-dimensional flats in $\mathrm{PG}(n,q)$ and $\mathrm{AG}(n,q)$ are $\begin{bmatrix} n+1 \\ k+1 \end{bmatrix}_q$ and $q^{n-k}\begin{bmatrix} n \\ k \end{bmatrix}_q$ respectively.

Projective and affine spaces provide important examples of designs, whose parameters can be expressed in terms of the Gaussian coefficients.

A *t-design* with *parameters* (v,k,λ), or *t-(v,k,λ) design*, consists of a set X of *v points*, and a collection $\mathcal{B}$ of k-element subsets of X called *blocks*, with the property that any t distinct points of X are contained in exactly λ blocks. Designs were first used by statisticians, such as R. A. Fisher, for experimental design (e.g. to facilitate analysis of variance). The terms "design" and "block", and the letter v (the initial letter of "variety"), reflect this origin.

Proposition 1.13 *(a) The points and m-dimensional flats in $\mathrm{PG}(n,q)$ form a 2-design with parameters*

$$\left(\begin{bmatrix} n+1 \\ 1 \end{bmatrix}_q, \begin{bmatrix} m+1 \\ 1 \end{bmatrix}_q, \begin{bmatrix} n-1 \\ m-1 \end{bmatrix}_q \right).$$

(b) The points and m-dimensional flats of $\mathrm{AG}(n,q)$ form a 2-design with parameters

$$\left(q^n,\, q^m,\, \begin{bmatrix} n-1 \\ m-1 \end{bmatrix}_q \right).$$

If $q = 2$, then it is a 3-design, with $\lambda = \begin{bmatrix} n-2 \\ m-2 \end{bmatrix}_2$.

Proof The values of v and k are clear in both cases.

(a) Let V be the underlying vector space of rank $n+1$. We want to count the subspaces of rank $m+1$ containing two given rank 1 subspaces P_1 and P_2. If $L = P_1 + P_2$, then L has rank 2, and a subspace contains P_1 and P_2 if and only if it contains L. Now, by the Third Isomorphism Theorem, the rank $m+1$ subspaces containing L are in 1-1 correspondence with the rank $m-1$ subspaces of the rank $n-1$ space V/L.

(b) In $\mathrm{AG}(n,q)$, to count subspaces containing two points, we may assume (by translation) that one of the points is the origin. An affine flat containing the origin is a vector subspace, and a subspace contains a non-zero vector if and only if it contains the rank 1 subspace it spans. The result follows as before. In the case when $q = 2$, a rank 1 subspace contains only one non-zero vector, so any two distinct non-zero vectors span a rank 2 subspace. ∎

Remark The essence of the proof is that the quotient of either $\mathrm{PG}(n,q)$ or $\mathrm{AG}(n,q)$ by a flat F of dimension d is $\mathrm{PG}(n-d-1,q)$. (The flats of the quotient space are precisely the flats of the original space containing F.) This assertion is true over any field at all, and lies at the basis of an approach to geometry which we will consider in Chapter 5.

An *automorphism* of a design is a permutation of the points which maps any block to a block.

Proposition 1.14 *For $0 < m < n$, the design of points and m-dimensional flats in $\mathrm{PG}(n,q)$ or $\mathrm{AG}(n,q)$ is $\mathrm{P\Gamma L}(n+1,q)$ or $\mathrm{A\Gamma L}(n+1,q)$ respectively, except in the affine case with $q=2$ and $m=1$.*

Proof By the results of Section 1.3, it suffices to show that the entire geometry can be recovered from the points and m-dimensional flats. This follows immediately from two observations:

(a) the unique line containing two points is the intersection of all the m-dimensional flats containing them;

(b) except for affine spaces over $\mathrm{GF}(2)$, a set of points is a flat if and only if it contains the line through any two of its points.

Affine spaces over $\mathrm{GF}(2)$ are exceptional: lines have just two points, and any two points form a line. However, analogous statements hold for planes: three points lie in a unique plane, and we have

(aa) the plane through three points is the intersection of all the flats of dimension m which contain them (for $m > 1$);

(bb) a set of points is a flat if and only if it contains the plane through any three of its points.

The proofs are left as exercises. ■

Exercises

1. Prove the assertions (a), (b), (aa), (bb) in Proposition 1.14.

2. Prove that the probability that a random $n \times n$ matrix over a given finite field $\mathrm{GF}(q)$ is non-singular tends to a limit $c(q)$ as $n \to \infty$, where $0 < c(q) < 1$.

3. Prove that the total number $F(n)$ of subspaces of a vector space of rank n over a given finite field $\mathrm{GF}(q)$ satisfies the recurrence

$$F(n+1) = 2F(n) + (q^n - 1)F(n-1).$$

4. Let S be an "alphabet" of size q, with two distinguished elements 0 and 1 (but not necessarily a finite field). A $k \times n$ matrix with entries from S is (as usual) in *reduced echelon form* if

- it has no zero rows;
- the first non-zero entry in any row is a 1;
- the "leading 1s" in later rows occur further to the right;
- the other entries in the column of a "leading 1" are all 0.

Prove that the number of $k \times n$ matrices in reduced echelon form is $\begin{bmatrix} n \\ k \end{bmatrix}_q$. Verify in detail in the case $n = 4$, $k = 2$.

5. Use the result of Exercise 4 to prove the recurrence relation

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = q^n \begin{bmatrix} n-1 \\ k \end{bmatrix}_q + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q.$$

2

Projective planes

Projective and affine planes are more than just spaces of smallest (non-trivial) dimension: as we will see, they are truly exceptional, and also they play a crucial rôle in the coordinatisation of arbitrary spaces.

2.1 Projective planes

We have seen in Sections 1.2 and 1.3 that, for any field F, the geometry $\mathrm{PG}(2,F)$ has the following properties:

(PP1) Any two points lie on exactly one line.

(PP2) Any two lines meet in exactly one point.

(PP3) There exist four points, no three of which are collinear.

I will now use the term *projective plane* in a more general sense, to refer to any structure of points and lines which satisfies conditions (PP1)-(PP3) above.

In a projective plane, let p and L be a point and line which are not incident. The incidence defines a bijection between the points on L and the lines through p. By (PP3), given any two lines, there is a point incident with neither; so the two lines contain equally many points. Similarly, each point lies on the same number of lines; and these two constants are equal. The *order* of the plane is defined to be one less than this number. The order of $\mathrm{PG}(2,F)$ is equal to the cardinality of F. (We saw in the last section that a projective line over $\mathrm{GF}(q)$ has $\begin{bmatrix}2\\1\end{bmatrix}_q = q+1$ points; so $\mathrm{PG}(2,q)$ is a projective plane of order q. In the infinite case, the claim follows by simple cardinal arithmetic.)

Given a finite projective plane of order n, each of the $n+1$ lines through a point p contains n further points, with no duplications, and all points are accounted for in this way. So there are n^2+n+1 points, and the same number of lines. The points and lines form a 2-$(n^2+n+1, n+1, 1)$ design. The converse is also true (see Exercise 2).

Do there exist projective planes not of the form PG$(2,F)$? The easiest such examples are infinite; I give two completely different ones below. Finite examples will appear later.

Example 1: *Free planes.* Start with any configuration of points and lines having the property that two points lie on at most one line (and dually), and satisfying (PP3). Perform the following construction. At odd-numbered stages, introduce a new line incident with each pair of points not already incident with a line. At even-numbered stages, act dually: add a new point incident with each pair of lines for which such a point doesn't yet exist. After countably many stages, a projective plane is obtained. For given any two points, there will be an earlier stage at which both are introduced; by the next stage, a unique line is incident with both; and no further line incident with both is added subsequently; so (PP1) holds. Dually, (PP2) holds. Finally, (PP3) is true initially and remains so. If we start with a configuration violating Desargues' Theorem (for example, the Desargues configuration with the line pqr "broken" into separate lines pq, qr, rp), then the resulting plane doesn't satisfy Desargues' Theorem, and so is not a PG$(2,F)$.

Example 2: *Moulton planes.* Take the ordinary real affine plane. Imagine that the lower half-plane is a refracting medium which bends lines of positive slope so that the part below the axis has twice the slope of the part above, while lines with negative (or zero or infinite) slope are unaffected. This is an affine plane, and has a unique completion to a projective plane (see later). The resulting projective plane fails Desargues' theorem. To see this, draw a Desargues configuration in the ordinary plane in such a way that just one of its ten points lies below the axis, and just one line through this point has positive slope.

The first examples of finite planes in which Desargues' Theorem fails were constructed by Veblen and Wedderburn [38]. Many others have been found since, but all known examples have prime power order. The *Bruck–Ryser Theorem* [4] asserts that, if a projective plane of order n exists, where $n \equiv 1$ or 2 (mod 4), then n must be the sum of two squares. Thus, for example, there is no projective plane of order 6 or 14. This theorem gives no information about 10, 12, 15, 18, ...

Recently, Lam, Swiercz and Thiel [21] showed by an extensive computation that there is no projective plane of order 10. The other values mentioned are undecided.

An *affine plane* is an incidence structure of points and lines satisfying the following conditions (in which two lines are called *parallel* if they are equal or disjoint):

(AP1) Two points lie on a unique line.

(AP2) Given a point p and line L, there is a unique line which contains p and is parallel to L.

(AP3) There exist three non-collinear points.

Remark. Axiom (AP2) for the real plane is an equivalent form of Euclid's "parallel postulate". It is called "Playfair's Axiom", although it was stated explicitly by Proclus.

Again it holds that $\mathrm{AG}(2,F)$ is an affine plane. More generally, if a line and all its points are removed from a projective plane, the result is an affine plane. (The removed points and line are said to be "at infinity". Two lines are parallel if and only if they contain the same point at infinity.

Conversely, let an affine plane be given, with point set $\mathcal{P}$ and line set $\mathcal{L}$. It follows from (AP2) that parallelism is an equivalence relation on $\mathcal{L}$. Let $\mathcal{Q}$ be the set of equivalence classes. For each line $L \in \mathcal{L}$, let $L^+ = L \cup \{Q\}$, where Q is the parallel class containing L. Then the structure with point set $\mathcal{P} \cup \mathcal{Q}$, and line set $\{L^+ : L \in \mathcal{L}\} \cup \{\mathcal{Q}\}$, is a projective plane. Choosing $\mathcal{Q}$ as the line at infinity, we recover the original affine plane.

We will have more to say about affine planes in Section 3.5.

Exercises

1. Show that a structure which satisfies (PP1) and (PP2) but not (PP3) must be of one of the following types:

(a) There is a line incident with all points. Any further line is a singleton, repeated an arbitrary number of times.

(b) There is a line incident with all points except one. The remaining lines all contain two points, the omitted point and one of the others.

2. Show that a 2-$(n^2+n+1, n+1, 1)$ design (with $n > 1$) is a projective plane of order n.

3. Show that, in a finite affine plane, there is an integer $n > 1$ such that

- every line has n points;
- every point lies on $n+1$ lines;
- there are n^2 points;
- there are $n+1$ parallel classes with n lines in each.

(The number n is the *order* of the affine plane.)

4. (The *Friendship Theorem*.) In a finite society, any two individuals have a unique common friend. Prove that there exists someone who is everyone else's friend.

[Let X be the set of individuals, $\mathcal{L} = \{F(x) : x \in X\}$, where $F(x)$ is the set of friends of X. Prove that, in any counterexample to the theorem, $(X, \mathcal{L})$ is a projective plane, of order n, say.

Now let A be the real matrix of order n^2+n+1, with (x,y) entry 1 if x and y are friends, 0 otherwise. Prove that

$$A^2 = nI + J,$$

where I is the identity matrix and J the all-1 matrix. Hence show that the real symmetric matrix A has eigenvalues $n+1$ (with multiplicity 1) and $\pm\sqrt{n}$. Using the fact that A has trace 0, calculate the multiplicity of the eigenvalue $\sqrt{n}$, and hence show that $n = 1$.]

5. Show that any Desargues configuration in a free projective plane must lie within the starting configuration. [Hint: Suppose not, and consider the last point or line to be added.]

2.2 Desarguesian and Pappian planes

It is no coincidence that we distinguished the free and Moulton planes from PG$(2,F)$s in the last section by the failure of Desargues' Theorem.

Theorem 2.1 *A projective plane is isomorphic to* PG$(2,F)$ *for some F if and only if it satisfies Desargues' Theorem.*

I do not propose to give a detailed proof of this important result; but some comments on the proof are in order.

We saw in Section 1.3 that, in PG$(2,F)$, the field operations (addition and multiplication) can be defined geometrically, once a set of four points with no three

collinear has been chosen. By (PP3), such a set of points exists in any projective plane. So it is possible to define two binary operations on a set consisting of a line with a point removed, and to coordinatise the plane with this algebraic object. Now it is obvious that any field axiom translates into a certain "configuration theorem", so that the plane is a $\mathrm{PG}(2,F)$ if and only if all these "configuration theorems" hold. What is not obvious, and quite remarkable, is that all these "configuration theorems" follow from Desargues' Theorem.

Another method, more difficult in principle but much easier in detail, exploits the relation between Desargues' Theorem and collineations.

Let p be a point and L a line. A *central collineation* with centre p and axis L is a collineation fixing every point on L and every line through p. It is called an *elation* if p is on L, a *homology* otherwise. The central collineations with centre p and axis L form a group. The plane is said to be (p,L)*-transitive* if this group permutes transitively the set $M \setminus \{p, L \cap M\}$ for any line $M \neq L$ on p (or, equivalently, the set of lines on q different from L and pq, where $q \neq p$ is a point of L).

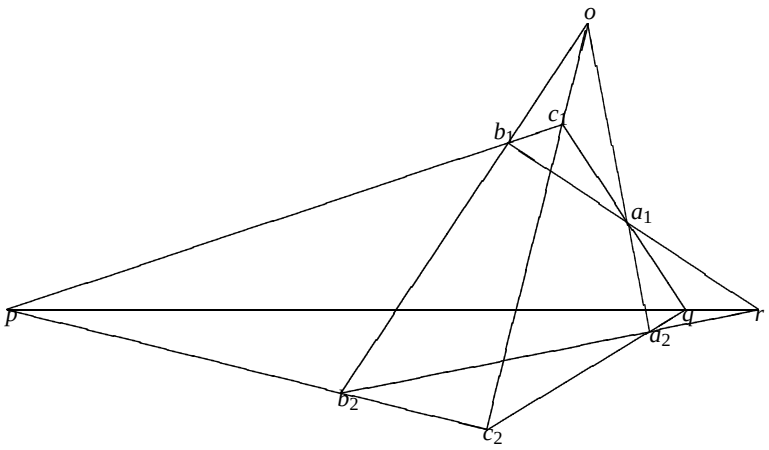


Figure 2.1: The Desargues configuration

Theorem 2.2 *A projective plane satisfies Desargues' Theorem if and only if it is* (p,L)*-transitive for all points* p *and lines* L*.*

Proof Let us take another look at the Desargues configuration (Fig. 2.1). It is clear that any central configuration with centre o and axis L which carries a_1 to a_2 is completely determined at every point b_1 not on M. (The line a_1a_2 meets L at a fixed point r and is mapped to b_1b_2; so b_2 is the intersection of ra_2 and ob_1.) Now, if we replace M with another line M' through o, we get another determination of the action of the collineation. It is easy to see that the condition that these two specifications agree is precisely Desargues' Theorem.

The proof shows a little more. Once the action of the central collineation on one point of $M \setminus \{o, L \cap M\}$ is known, the collineation is completely determined. So, if Desargues' Theorem holds, then these groups of central collineations act sharply transitively on the relevant set.

Now the additive and multiplicative structures of the field turn up as groups of elations and homologies respectively with fixed centre and axis. We see immediately that these structures are both groups. More of the axioms are easily deduced too. For example, let L be a line, and consider all elations with axis L (and arbitrary centre on L). This set is a group G. For each point p on L, the elations with centre p form a normal subgroup. These normal subgroups partition the non-identity elements of G, since a non-identity elation has at most one centre. But a group having such a partition is abelian (see Exercise 2). So addition is commutative.

In view of this theorem, projective planes over skew fields are called *Desarguesian planes*.

There is much more to be said about the relationships among configuration theorems, coordinatisation, and central collineations. I refer to Dembowski's book for some of these. One such relation is of particular importance.

Pappus' Theorem is the assertion that, if alternate vertices of a hexagon are collinear (that is, the first, third and fifth, and also the second, fourth and sixth), then also the three points of intersection of opposite edges are collinear. See Fig. 2.2.

Theorem 2.3 *A projective plane satisfies Pappus' Theorem if and only if it is isomorphic to* $\mathrm{PG}(2,F)$ *for some commutative field* F.

Proof The proof involves two steps. First, a purely geometric argument shows that Pappus' Theorem implies Desargues'. This is shown in Fig. 2.3. This figure shows a potential Desargues configuration, in which the required collinearity is shown by three applications of Pappus' Theorem. The proof requires four new

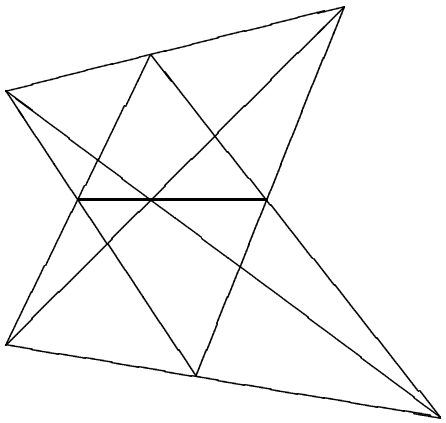


Figure 2.2: Pappus' Theorem

points, $s = a_1b_1 \cap a_2c_2$, $t = b_1c_1 \cap os$, $u = b_1c_2 \cap oa_1$, and $v = b_2c_2 \cap os$. Now Pappus' Theorem, applied to the hexagon $osc_2b_1c_1a_1$, shows that q, u, t are collinear; applied to $osb_1c_2b_2a_2$, shows that r, u, v are collinear; and applied to b_1tuvc_2s (using the two collinearities just established), shows that p, q, r are collinear. The derived collinearities are shown as dotted lines in the figure. (Note that the figure shows only the generic case of Desargues' Theorem; it is necessary to take care of the possible degeneracies as well.)

The second step involves the use of coordinates to show that, in a Desarguesian plane, Pappus' Theorem is equivalent to the commutativity of multiplication. (See Exercise 3.)

In view of this, projective planes over commutative fields are called *Pappian planes*.

Remark. It follows from Theorems 2.1 and 2.3 and Wedderburn's Theorem 1.1 that, in a finite projective plane, Desargues' Theorem implies Pappus'. No geometric proof of this implication is known.

A similar treatment of affine planes is possible.

Exercises

1. (a) Show that a collineation which has a centre has an axis, and *vice versa*.

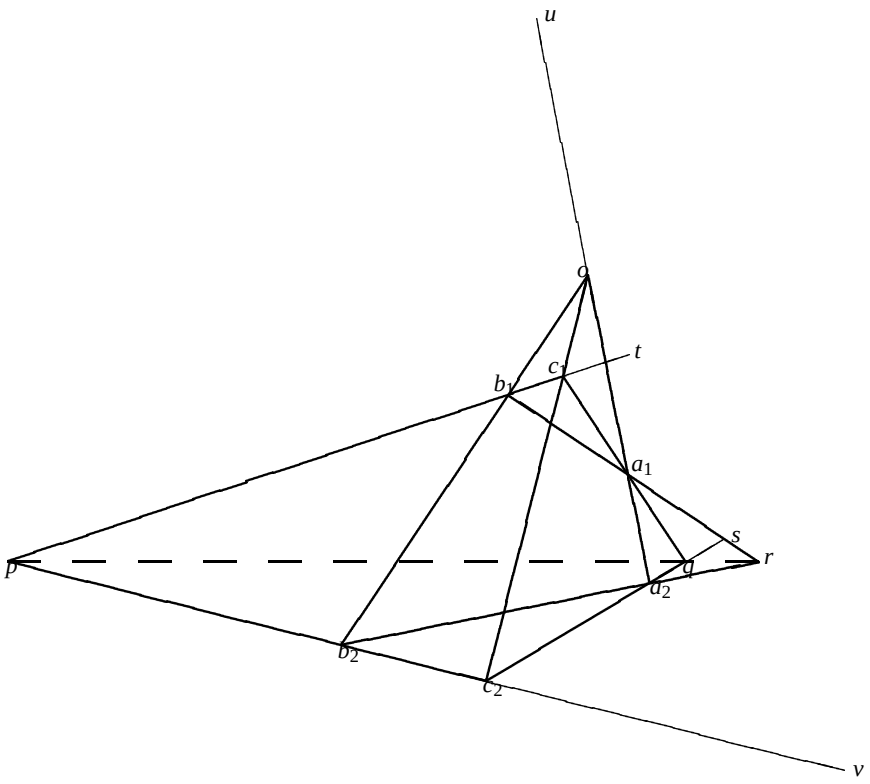


Figure 2.3: Pappus implies Desargues

(b) Show that a collineation cannot have more than one centre.

2. The group G has a family of proper normal subgroups which partition the non-identity elements of G. Prove that G is abelian.

3. In PG$(2,F)$, let the vertices of a hexagon be $(1,0,0)$, $(0,0,1)$, $(0,1,0)$, $(1,\alpha+1,1)$, $(1,1,0)$ and $(\beta,\beta(\alpha+1),1)$. Show that alternate vertices lie on the lines defined by the column vectors $(0,0,1)^\top$ and $(\alpha+1,-1,0)^\top$. Show that opposite sides meet in the points $(\alpha,0,-1)$, $(0,\beta\alpha,1)$ and $(1,\beta(\alpha+1),1)$. Show that the second and third of these lie on the line $(\beta,-1,\beta\alpha)^\top$, which also contains the first if and only if $\alpha\beta=\beta\alpha$.

2.3 Projectivities

Let $\Pi = (X, \mathcal{L})$ be a projective plane. Temporarily, let (L) be the set of points incident with L; and let (x) be the set of lines incident with x. If x is not incident with L, there is a natural bijection between (L) and (x): each point on L lies on a unique line through x. This bijection is called a *perspectivity*. By iterating perspectivities and their inverses, we get a bijection (called a *projectivity*) between any two sets (x) or (L). In particular, for any line L, we obtain a set $P(L)$ of projectivities from (L) to itself (or *self-projectivities*), and analogously a set $P(x)$ for any point x.

The sets $P(L)$ and $P(x)$ are actually groups of permutations of (L) or (x). (Any self-projectivity is the composition of a chain of perspectivities; the product of two self-projectivities corresponds to the concatenation of the chains, while the inverse corresponds to the chain in reverse order.) Moreover, these permutation groups are naturally isomorphic: if g is any projectivity from (L_1) to (L_2), say, then $g^{-1}P(L_1)g = P(L_2)$. So the group $P(L)$ of self-projectivities on a line is an invariant of the projective plane. It turns out that the structure of this group carries information about the plane which is closely related to concepts we have already seen.

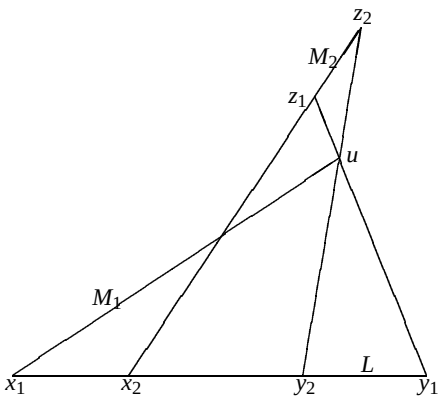


Figure 2.4: 3-transitivity

Proposition 2.4 *The permutation group $P(L)$ is 3-transitive.*

Proof It suffices to show that there is a projectivity fixing any two points $x_1, x_2 \in L$ and mapping any further point y_1 to any other point y_2. In general, we will use

the notation "(L_1 to L_2 via p)" for the composite of the perspectivities $(L_1) \to (p)$ and $(p) \to (L_2)$. Let M_i be any other lines through x_i ($i = 1,2$), u a point on M_1, and $z_i \in M_2$ ($i = 1,2$) such that $y_i u z_i$ are collinear ($i = 1,2$). Then the product of (L to M_1 via z_1) and (M_1 to L via z_2) is the required projectivity (Fig. 2.4.)

A permutation group G is *sharply t-transitive* if, given any two t-tuples of distinct points, there is a unique element of G carrying the first to the second (in order). The main result about groups of projectivities is the following theorem:

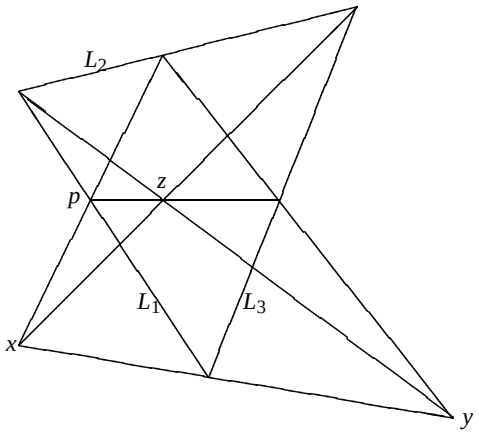


Figure 2.5: Composition of projectivities

Theorem 2.5 *The group $P(L)$ of projectivities on a projective plane Π is sharply 3-transitive if and only if Π is pappian.*

Proof We sketch the proof. The crucial step is the equivalence of Pappus' Theorem to the following assertion:

> Let L_1, L_2, L_3 be non-concurrent lines, and x and y two points such that the projectivity
>
> $$g = (L_1 \text{ to } L_2 \text{ via } x) \cdot (L_2 \text{ to } L_3 \text{ via } y)$$
>
> fixes $L_1 \cap L_3$. Then there is a point z such that the projectivity g is equal to (L_1 to L_3 via z).

The hypothesis is equivalent to the assertion that x, y and $L_1 \cap L_3$ are collinear. Now the point z is determined, and Pappus' Theorem is equivalent to the assertion that it maps a random point p of L_1 correctly. (Fig. 2.5 is just Pappus' Theorem.)

Now this assertion allows long chains of projectivities to be shortened, so that their action can be controlled.

The converse can be seen another way. By Theorem 2.3, we know that a Pappian plane is isomorphic to $\mathrm{PG}(2,F)$ for some commutative field F. Now it is easily checked that any self-projectivity on a line is induced by a linear fractional transformation (an element of $\mathrm{PGL}(2,F)$; and this group is sharply 3-transitive.

In the finite case, there are very few 3-transitive groups apart from the symmetric and alternating groups; and, for all known non-Pappian planes, the group of projectivities is indeed symmetric or alternating (though it is not known whether this is necessarily so). Both possibilities occur; so, at present, all that this provides us for non-Pappian finite planes is a single Boolean invariant.

In the infinite case, however, more interesting possibilities arise. If the plane has order α, then the group of projectivities has α generators, and so has order α; so it can never be the symmetric group (which has order 2^α). Barlotti [1] gave an example in which the stabiliser of any six points is the identity, and the stabiliser of any five points is a free group. On the other hand, Schleiermacher [25] showed that, if the stabiliser of any five points is trivial, then the stabiliser of any three points is trivial (and the plane is Pappian).

Further developments involve deeper relationships between projectivities, configuration theorems, and central collineations; the definition and study of projectivities in other incidence structures; and so on.

3

Coordinatisation of projective spaces

In this chapter, we describe axiom systems for projective (and affine) spaces. The principal results are due to Veblen and Young.

3.1 The $\mathrm{GF}(2)$ case

In the last section, we saw an axiomatic characterisation of the geometries $\mathrm{PG}(2,F)$ (as projective planes satisfying Desargues' Theorem). We turn now to the characterisation of projective spaces of arbitrary dimension, due to Veblen and Young. Since the points and the subspaces of any fixed dimension determine the geometry, we expect an axiomatisation in terms of these. Obviously the case of points and lines will be the simplest.

For the first of several times in these notes, we will give a detailed and self-contained argument for the case of $\mathrm{GF}(2)$, and treat the general case in rather less detail.

Theorem 3.1 *Let X be a set of points, $\mathcal{L}$ a set of subsets of X called lines. Assume:*

(a) any two points lie on a unique line;

(b) a line meeting two sides of a triangle, not at a vertex, meets the third side;

(c) a line contains exactly three points.

Then X and $\mathcal{L}$ are the sets of points and lines in a (not necessarily finite dimensional) projective space over $\mathrm{GF}(2)$.

htb

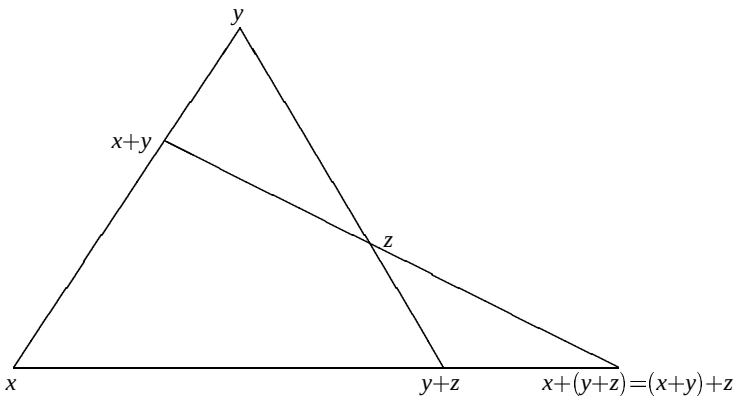


Figure 3.1: Veblen's Axiom

Remark We will see later that, more or less, conditions (a) and (b) characterise arbitrary projective spaces. Condition (c) obviously specifies that the field is GF(2). The phrase "not necessarily finite dimensional" should be interpreted as meaning that X and $\mathcal{L}$ can be identified with the subspaces of rank 1 and 2 respectively of a vector space over GF(2), not necessarily of finite rank.

Proof Since 1 is the only non-zero scalar in GF(2), the points of projective space can be identified with the non-zero vectors; lines are then triples of non-zero vectors with sum 0. Our job is to reconstruct this space.

Let 0 be an element not in X, and set $V = X \cup \{0\}$. Now define an addition in V as follows:

- for all $v \in V$, $0+v = v+0 = v$ and $v+v = 0$;
- for all $x,y \in X$ with $x \neq y$, $x+y = z$, where z is the third point of the line containing x and y.

We claim that $(V,+)$ is an abelian group. Commutativity is clear; 0 is the identity, and each element is its own inverse. Only the associative law is non-trivial; and the only non-trivial case, when x,y,z are distinct non-collinear points, follows immediately from Veblen's axiom (b) (see Fig. 3.1.1).

Next, we define scalar multiplication over GF(2), in the only possible way: $0 \cdot v = 0$, $1 \cdot v = v$ for all $v \in V$. The only non-trivial vector space axiom is $(1+1) \cdot v = 1 \cdot v + 1 \cdot v$, and this follows from $v+v = 0$.

Finally, $\{0,x,y,z\}$ is a rank 2 subspace if and only if $x+y = z$.

There is a different but even simpler characterisation in terms of hyperplanes, which foreshadows some later developments.

Let $\mathcal{L}$ be any family of subsets of X. The subset Y of X is called a *subspace* if any member of $\mathcal{L}$ which contains two points of Y is wholly contained within Y. (Thus, the empty set, the whole of X, and any singleton are trivially subspaces.) The subspace Y is called a *hyperplane* if it intersects every member of $\mathcal{L}$ (necessarily in one or all of its points).

Theorem 3.2 *Let $\mathcal{L}$ be a collection of subsets of X. Suppose that*

(a) every set in $\mathcal{L}$ has cardinality 3;

(b) any two points of X lie in at least one member of $\mathcal{L}$;

(c) every point of X lies outside some hyperplane.

Then X and $\mathcal{L}$ are the point and line sets of a projective geometry over GF(2)*, not necessarily finite dimensional.*

Proof Let $\mathcal{H}$ be the set of hyperplanes. For each point $x \in X$, we define a function $p_x : \mathcal{H} \to \mathrm{GF}(2)$ by the rule

$$p_x(H) = \begin{cases} 0 & \text{if } x \in H; \\ 1 & \text{if } x \notin H. \end{cases}$$

By condition (c), p_x is non-zero for all $x \in X$.

Let $P = \{p_x : x \in X\}$. We claim that $P \cup \{0\}$ is a subspace of the vector space $\mathrm{GF}(2)^{\mathcal{H}}$ of functions from $\mathcal{H}$ to GF(2). Take $x, y \in X$, and let $\{x, y, z\}$ be any set in $\mathcal{L}$ containing x and y. Then a hyperplane contains z if and only if it contains both or neither of x and y; so $p_z = p_x + p_y$. The claim follows.

Now the map $x \mapsto p_x$ is 1–1, since if $p_x = p_y$ then $p_x + p_y = 0$, contradicting the preceding paragraph. Clearly this map takes members of $\mathcal{L}$ to lines. The theorem is proved.

Remark The fact that two points lie in a unique line turns out to be a consequence of the other assumptions.

Exercises

1. Suppose that conditions (a) and (c) of Theorem 3.1.2 hold. Prove that two points of X lie in at most one member of $\mathcal{L}$.

2. Let $(X, \mathcal{L})$ satisfy conditions (a) and (c) of Theorem 3.1.1. Let Y be the set of points x of X with the following property: for any two lines $\{x, y_1, y_2\}$ and $\{x, z_1, z_2\}$ containing x, the lines y_1z_1 and y_2z_2 intersect. Prove that Y is a subspace of X.

3. Let X be a set of points, $\mathcal{B}$ a collection of subsets of X called lines. Assume that any two points lie in at least one line, and that every point lies outside some hyperplane. Show that, if the line size is not restricted to be 3, then we cannot conclude that X and $\mathcal{B}$ are the point and line sets of a projective space, even if any two points lie on exactly one line. [Hint: In a projective plane, any line is a hyperplane. Select three lines L_1, L_2, L_3 forming a triangle. Show that it is possible to delete some points, and to add some lines, so that L_1, L_2, L_3 remain hyperplanes.]

4. Let $(X, \mathcal{B})$ satisfy the hypotheses of the previous question. Assume additionally that any two points lie on a unique line, and that some hyperplane is a line and is finite. Prove that there is a number n such that any hyperplane contains $n+1$ points, any point lies on $n+1$ lines, and the total number of lines is n^2+n+1.

3.2 An application

I now give a brief application to coding theory. This application is a bit spurious, since a more general result can be proved by a different but equally simple argument; but it demonstrates an important link between these fields. Additionally, the procedure can be reversed, to give characterisations of other combinatorial designs using theorems about codes.

The problem tackled by the theory of error-correcting codes is to send a message over a noisy channel in which some distortion may occur, so that the errors can be corrected but the price paid in loss of speed is not too great. This is not the place to discuss coding theory in detail. We simplify by assuming that a message transmitted over the channel is a sequence of blocks, each block being an n-tuple of *bits* (zeros or ones). We also assume that we can be confident that, during the transmission of a single block, no more than e bits are transmitted incorrectly (a zero changed to a one or *vice versa*). The *Hamming distance* between two blocks is the number of coordinates in which they differ; that is, the number of errors required to change one into the other. A *code* is just a set of "codewords" or blocks (n-tuples of bits), containing more than one codeword. It is *e-error correcting* if the Hamming distance between two codewords is at least $2e+1$. (The reason for the name is that, by the triangle inequality, an arbitrary word cannot lie at distance

e or less from more than one codeword. By our assumption, the received word lies at distance e or less from the transmitted codeword; so this codeword can be recovered.)

To maximise the transmission rate, we need as many codewords as possible. The optimum is obtained when every word lies within distance e of a (unique) codeword. In other words, the closed balls of radius e centred at the codewords fill the space of all words without any overlap! A code with this property is called *perfect e-error-correcting*.

Encoding and decoding are made much easier if the code is *linear*, that is, it is a $\mathrm{GF}(2)$-subspace of the vector space $\mathrm{GF}(2)^n$ of all words.

Theorem 3.3 *A linear perfect* 1*-error-correcting code has length* $2^d - 1$ *for some* $d > 1$*; there is a unique such code of any length having this form.*

Remark These unique codes are called *Hamming codes*. Their relation to projective spaces will be made clear by the proof below.

Proof Let C be such a code, of length n. Obviously it contains $\mathbf{0}$. We define the *weight* $\mathrm{wt}(\mathbf{v})$ of any word $\mathbf{v}$ to be its Hamming distance from $\mathbf{0}$. The weight of any non-zero codeword is at least 3. Now let X be the set of coordinate places, and $\mathcal{L}$ the set of triples of points of X which support codewords (i.e., for which a codeword has 1s in just those positions).

We verify the hypotheses of Theorem 3.1. Condition (c) is clear.

Let x and y be coordinate positions, and let $\mathbf{w}$ be the word with entries 1 in positions x and y and 0 elsewhere. $\mathbf{w}$ is not a codeword, so there must be a unique codeword $\mathbf{c}$ at distance 1 from $\mathbf{w}$; then $\mathbf{c}$ must have weight 3 and support containing x and y. So (a) holds.

Let $\{x,y,r\},\{x,z,q\},\{y,z,p\}$ be the supports of codewords $\mathbf{u},\mathbf{v},\mathbf{w}$. By linearity, $\mathbf{u}+\mathbf{v}+\mathbf{w}$ is a codeword, and its support is $\{p,q,r\}$. So (b) holds.

Thus X and $\mathcal{L}$ are the points and lines of a projective space $\mathrm{PG}(d-1,2)$ for some $d > 1$; the number of points is $n = 2^d - 1$. Moreover, it's easy to see that C is spanned by its words of weight 3 (see Exercise 1), so it is uniquely determined by d.

Note, incidentally, that the automorphism group of the Hamming code is the same as that of the projective space, viz. $\mathrm{PGL}(d,2)$.

Exercise

1. Prove that a perfect linear code is spanned by its words of minimum weight. (Use induction on the weight. If $\mathbf{w}$ is any non-zero codeword, there is a codeword $\mathbf{u}$ whose support contains $e+1$ points of the support of $\mathbf{w}$; then $\mathbf{u}+\mathbf{w}$ has smaller weight than $\mathbf{w}$.)

2. Prove that if a perfect e-error-correcting code of length n exists, then

$$\sum_{i=0}^{e} \binom{n}{i}$$

is a power of 2. Deduce that, if $e = 3$, then $n = 7$ or 23. (Hint: the cubic polynomial in n factorises.)

Remark. The case $n = 7$ is trivial. For $n = 23$, there is a unique code (up to isometry), the so-called *binary Golay code*.

3. Verify the following decoding scheme for the Hamming code H_d of length $2^d - 1$. Let M_d be the $2^d - 1 \times d$ matrix over GF(2) whose rows are the base 2 representations of the integers $1, 2, \ldots, 2^d - 1$. Show that the null space of the matrix M_d is precisely H_d. Now let $\mathbf{w}$ be received when a codeword is transmitted, and assume that at most one error has occurred. Prove that

- if $\mathbf{w}H_d = 0$, then $\mathbf{w}$ is correct;
- if $\mathbf{w}H_d$ is the i^{th} row of H_d, then the i^{th} position is incorrect.

3.3 The general case

The general coordinatisation theorem is the same as Theorem 3.1, with the hypothesis "three points per line" weakened to "at least three points per line". Accordingly we consider geometries with point set X and line set $\mathcal{L}$ (where $\mathcal{L}$ is a set of subsets of X) satisfying:

(LS1) Any line contains at least two points.

(LS2) Two points lie in a unique line.

Such a geometry is called a *linear space*. Recall that a subspace is a set of points which contains the (unique) line through any two of its points. In a linear space, in addition to the trivial subspaces (the empty set, singletons, and X), any line is

a subspace. Any subspace, equipped with the lines it contains, is a linear space in its own right.

A linear space is called *thick* if it satisfies:

(LS1+) Any line contains at least three points.

Finally, we will impose Veblen's Axiom:

(V) A line meeting two sides of a triangle, not at a vertex, meets the third side also.

Theorem 3.4 (Veblen–Young Theorem) *Let $(X, \mathcal{L})$ be a linear space, which is thick and satisfies Veblen's Axiom (V). Then one of the following holds:*

(a) $X = \mathcal{L} = \emptyset$*;*

(b) $|X| = 1$, $\mathcal{L} = \emptyset$*;*

(c) $\mathcal{L} = \{X\}$, $|X| \geq 3$*;*

(d) $(X, \mathcal{L})$ *is a projective plane;*

(e) $(X, \mathcal{L})$ *is a projective space over a skew field, not necessarily of finite dimension.*

Remark It is common to restrict to finite-dimensional projective spaces by adding the additional hypothesis that any chain of subspaces has finite length.

Proof (outline) The key observation provides us with lots of subspaces.

Lemma 3.5 *Let $(X, \mathcal{L})$ be a linear space satisfying Veblen's axiom. Let Y be a subspace, and p a point not in Y; let Z be the union of the lines joining p to points of Y. Then Z is a subspace, and Y is a hyperplane in Z.*

Proof Let q and r be points of Z. There are several cases, of which the generic case is that where $q, r \notin Y$ and the lines pq and pr meet Y in distinct points s, t. By (V), the lines qr and st meet at a point u of Y. If v is another point of qr, then by (V) again, the line pv meets st at a point of Y; so $v \in Z$.

We write this subspace as $\langle Y, p\rangle$.

Now, if L is a line and p a point not in L, then $\langle L, p\rangle$ is a projective plane. (It is a subspace in which L is a hyperplane; all that has to be shown is that every line is a hyperplane, which follows once we show that $\langle L, p\rangle$ contains no proper subspace properly containing a line.)

The theorem is clearly true if there do not exist four non-coplanar points; so we may suppose that such points do exist.

We claim that Desargues' Theorem holds. To see this examine the geometric proof of Desargues' Theorem in Section 1.2; it is obvious for any non-planar configuration, and the planar case follows by several applications of the non-planar case. Now the same argument applies here.

It follows from Theorem 2.1 that every plane in our space can be coordinatised by a skew field.

To complete the proof, we have to show that the coordinatisation can be extended consistently to the whole space. For this, first one shows that the skew fields coordinatising all planes are the same: this can be proved for planes within a 3-dimensional subspace by means of central collineations, and the result extends by connectedness to all pairs of planes. The remainder of the argument involves careful book-keeping.

From this, we can find a classification of not necessarily thick linear spaces satisfying Veblen's axiom. The *sum* of a family of linear spaces is defined as follows. The point set is the disjoint union of the point sets of the constituent spaces. Lines are of two types:

(a) all lines of the constituent spaces;

(b) all pairs of points from different constituents.

It is clearly a linear space.

Theorem 3.6 *A linear space satisfying Veblen's axiom is the sum of linear spaces of types (b)–(e) in the conclusion of Theorem 3.4.*

Proof Let $(X, \mathcal{L})$ be such a space. Define a relation $\sim$ on X by the rule that $x \sim y$ if either $x = y$, or the line containing x and y is thick (has at least three points). We claim first that $\sim$ is an equivalence relation. Reflexivity and symmetry are clear; so assume that $x \sim y$ and $y \sim z$, where we may assume that x, y and z are all distinct. If these points are collinear, then $x \sim z$; so suppose not; let x_1 and z_1 be

further points on the lines xy and yz respectively. By (V), the line x_1z_1 contains a point of xz different from x and z, as required.

So X is the disjoint union of equivalence classes. We show next that any equivalence class is a subspace. So let $x \sim y$. Then $x \sim z$ for every point z of the line xy; so this line is contained in the equivalence class of x.

So each equivalence class is a non-empty thick linear space, and hence a point, line, projective plane, or projective space over a skew field, by Theorem 3.4. It is clear that the whole space is the sum of its components.

A geometry satisfying the conclusion of Theorem 3.6 is called a *generalised projective space*. Its flats are its (linear) subspaces; these are precisely the sums of flats of the components. The term "projective space" is sometimes extended to mean "thick generalised projective space" (i.e., to include single points, lines with at least three points, and not necessarily Desarguesian projective planes).

3.4 Lattices

Another point of view is to regard the flats of a projective space as forming a lattice. We discuss this in the present section.

A *lattice* is a set L with two binary operations $\vee$ and $\wedge$ (called *join* and *meet*), and two constants 0 and 1, satisfying the following axioms:

(L1) $\vee$ and $\wedge$ are idempotent, commutative, and associative;

(L2) $x \vee (x \wedge y) = x$ and $x \wedge (x \vee y) = x$;

(L3) $x \wedge 0 = x$, $x \vee 1 = x$.

It follows from these axioms that $x \wedge y = x$ holds if and only if $x \vee y = y$ holds. We write $x \leq y$ if these equivalent conditions hold. Then $(L, \leq)$ is a partially ordered set with greatest element 1 and least element 0; $x \vee y$ and $x \wedge y$ are the least upper bound and greatest lower bound of x and y respectively. Conversely, any partially ordered set in which least upper bounds and greatest lower bounds of all pairs of elements exist, and there is a least element and a greatest element, gives rise to a lattice.

In a lattice, an *atom* is a non-zero element a such that $a \wedge x = 0$ or a for any x; in other words, an element greater than zero but minimal subject to this. The lattice is called *atomic* if every element is a join of atoms.

A lattice is *modular* if it satisfies:

(M) If $x \leq z$, then $x \vee (y \wedge z) = (x \vee y) \wedge z$ for all y.

(Note that, if $x \leq z$, then $x \vee (y \wedge z) \leq (x \vee y) \wedge z$ in any lattice.)

Theorem 3.7 *A lattice is a generalised projective space of finite dimension if and only if it is atomic and modular.*

Proof The forward implication is an exercise. Suppose that the lattice L is atomic and modular. Let X be the set of atoms. Identify every element z of the lattice with the set $\{x \in X : x \leq z\}$. (This map is 1–1; it translates meets to intersections, and the lattice order to the inclusion order.)

Let x, y, z be atoms, and suppose that $z \leq x \vee y$. Then trivially $x \vee z \leq x \vee y$. Suppose that these two elements are unequal. Then $y \not\leq x \vee z$. Since y is an atom, $y \wedge (x \vee z) = 0$, and so $x \vee (y \wedge (x \vee z)) = x$. But $(x \vee y) \wedge (x \vee z) = x \vee z$, contradicting modularity. So $x \vee z = x \vee y$. Hence, if we define lines to be joins of pairs of atoms, it follows that two points lie in a unique line.

Now we demonstrate Veblen's axiom. Let u, v be points on $x \vee y$, $x \vee z$ respectively, where xyz is a triangle. Suppose that $(y \vee z) \wedge (u \vee v) = 0$. Then $y \vee u \vee v \geq z$, so $y \vee u \vee v \geq y \vee z$; in other words, $y \vee (u \vee v)) \wedge (y \vee z) = y \vee z$. On the other hand, $y \vee ((u \vee v) \wedge (y \vee z)) = y \vee 0 = y$, contradicting modularity. So the lines $y \vee z$ and $u \vee v$ meet.

By Theorem 3.6, the linear subspace is a generalised projective geometry. Clearly the geometry has finite dimension. We leave it as an exercise to show that every flat of the geometry is an element of the lattice.

Exercises

1. Complete the proof of Theorem 3.7.
2. Show that an atomic lattice satisfying the distributive laws is modular, and deduce that it is isomorphic to the lattice of subsets of a finite set.

3.5 Affine spaces

Veblen's axiom in a linear space is equivalent to the assertion that three non-collinear points lie in a subspace which is a projective plane. It might be hoped that replacing "projective plane" by "affine plane" here would give an axiomatisation of affine spaces. We will see that this is almost true.

Recall from Section 2.1 the definition of an affine plane, and the fact that parallelism is an equivalence relation in an affine plane, where two lines are parallel if they are equal or disjoint.

Now suppose that $(X, \mathcal{L})$ is a linear space satisfying the following condition:

(AS1) There is a collection $\mathcal{A}$ of subspaces with the properties that each member of $\mathcal{A}$ is an affine plane, and that any three non-collinear points are contained in a unique member of $\mathcal{A}$.

First, a few remarks about such spaces.

1. All lines have the same cardinality. For two intersecting lines lie in an affine plane, and so are equicardinal; and, given two disjoint lines, there is a line meeting both.

2. It would be simpler to say "any three points generate an affine plane", where the subspace *generated by* a set is the intersection of all subspaces containing it. This formulation is equivalent if the cardinality of a line is not 2. (Affine spaces of order greater than 2 have no non-trivial proper subspaces.) But, if lines have cardinality 2, then any pair of points is a line, and so any three points form a subspace which is a generalised projective plane. However, we do want a formulation which includes this case.

3. In a linear space satisfying (AS1), two lines are said to be *parallel* if either they are equal, or they are disjoint and contained in a member of $\mathcal{A}$ (and hence parallel there). Now Playfair's Axiom holds: given a line L and point p, there is a unique line parallel to L and containing p. Moreover, parallelism is reflexive and symmetric, but not necessarily transitive. We will impose the further condition:

(AS2) Parallelism is transitive.

Theorem 3.8 *A linear space satisfying (AS1) and (AS2) is empty, a single point, a single line, an affine plane, or the configuration of points and lines in a (not necessarily finite-dimensional) affine space.*

Proof Let $(X, \mathcal{L})$ be the linear space. We may assume that it is not empty, a point, a line, or an affine plane (i.e., that there exist four non-coplanar points).

Step 1. Define a *solid* to be the union of all the lines in a parallel class C which meet a plane $\Pi \in \mathcal{A}$, where Π contains no line of C. Then any four non-coplanar points lie in a unique solid, and any solid is a subspace.

That a solid is a subspace is shown by considering cases, of which the generic one runs as follows. Let p, q be points such that the lines of C containing p and q meet Π in distinct points x and y. Then x, y, p, q lie in an affine plane; so the line of C through a point r of pq meets Π in a point x of xy.

Now the fact that the solid is determined by any four non-coplanar points follows by showing that it has no non-trivial proper subspaces except planes (if the cardinality of a line is not 2) or by counting (otherwise).

In a solid, if a plane Π contains no parallel to a line L, then Π meets L in a single point. Hence any two planes in a solid are disjoint or meet in a line.

Step 2. If two planes Π and Π' contain lines from two different parallel classes, then every line of Π is parallel to a line of Π'.

Suppose not, and let L, M, N be lines of Π, concurrent at p, and p' a point of Π' such that the lines L', M' through p' parallel to L and M lie in Π', but the line N' parallel to N does not. The whole configuration lies in a solid; so the planes NN' and Π', with a common point p', meet in a line K. Now K is coplanar with N but not parallel to it, so $K \cap N$ is a point q. Then Π and Π' meet in q, and hence in a line J. But then J is parallel to both L and M, a contradiction.

We call two such planes *parallel*.

Step 3. We build the embedding projective space. Here I will use a typographic convention to distinguish the two related spaces: elements of the space we are building will be written in CAPITALS. The POINTS are the points of X and the parallel classes of lines of $\mathcal{A}$. The LINES are the lines of $\mathcal{L}$ and the parallel classes of planes in $\mathcal{A}$. Incidence is hopefully obvious: as in the old space, together with incidence between any line and its parallel class, as well as between a parallel class C of lines and a parallel class $\mathcal{C}$ of planes if a plane in $\mathcal{C}$ contains a line in C.

By Step 2, this is a linear space; and clearly every LINE contains at least three POINTS. We call the new POINTS and LINES (i.e., the parallel classes) "ideal".

Step 4. We verify Veblen's Axiom. Any three points which are not all "ideal" lie in an affine plane with its points at infinity adjoined, i.e., a projective plane. So let pqr be a triangle of "ideal" POINTS, s and t POINTS on pq and pr respectively, and o a point of X. Let P, Q, R, S, T be the lines through o in the parallel classes p, q, r, s, t respectively. Then these five lines lie in a solid, so the planes QR and ST (having the point o in common) meet in a line u. The parallel class U of u is the required POINT on qr and st.

By Theorem 3.4, the extended geometry is a projective space. The points at infinity obviously form a hyperplane, and so the original points and lines form an affine space.

We spell the result out in the case where lines have cardinality 2, but referring only to parallelism, not to the planes.

Corollary 3.9 *Suppose that the 2-element subsets of a set X are partitioned into "parallel classes" so that each class partitions X. Suppose that, for any four points $p,q,r,s \in X$, if $pq \parallel rs$, then $pr \parallel qs$. Then the points and parallelism are those of an affine space over* $\mathrm{GF}(2)$.

Here, we have used the notation $\parallel$ to mean "belong to the same parallel class as". The result follows immediately from the theorem, on defining $\mathcal{A}$ to be the set of 4-element subsets which are the union of two parallel 2-subsets.

Exercises

1. Give a direct proof of the Corollary, in the spirit of Section 3.1.

3.6 Transitivity of parallelism

A remarkable theorem of Buekenhout [6] shows that it is not necessary to assume axiom (AS2) (the transitivity of parallelism) in Theorem 3.8, provided that the cardinality of a line is at least 4. Examples due to Hall [19] show that the condition really is needed if lines have cardinality 3.

Theorem 3.10 *Let $(X,\mathcal{L})$ be a linear space satisfying (AS1), in which some line contains at least four points. Then parallelism is transitive (that is, (AS2) holds), and so $(X,\mathcal{L})$ is an affine space.*

To discuss the counterexamples with 2 or 3 points on a line, some terminology is helpful. A *Steiner triple system* is a collection of 3-subsets of a set, any two points lying in a unique subset of the collection. In other words, it is a linear space with all lines of cardinality 3, or (in the terminology of Section 1.4) a 2-$(v,3,1)$ design for some (possibly infinite) v. A *Steiner quadruple system* is a set of 4-subsets of a set, any three points in a unique subset in the collection (that is, a 3-$(v,4,1)$ design.)

A linear space satisfying (AS1), with two points per line, is equivalent to a Steiner quadruple system: the distinguished 4-sets are the affine planes. There are Steiner quadruple systems aplenty; most are not affine spaces over $\mathrm{GF}(2)$ (for example, because the number of points is not a power of 2). Here is an example. Let $A = \{1,2,3,4,5,6\}$. Let X be the set of all partitions of A into two sets of size 3 (so that $|X| = 10$). Define two types of 4-subsets of X:

(a) for all $a, b \in A$, the set of partitions for which a, b lie in the same part;

(b) for all partitions of A into three 2-sets A_1, A_2, A_3, the set of all partitions into two 3-sets each of which is a transversal to the three sets A_i.

This is a Steiner quadruple system with 10 points.

In the case of three points per line, we have the following result, for which we refer to Bruck [D] and Hall [18, 19]:

Theorem 3.11 *(a) In a finite Steiner triple system satisfying (AS1), the number of points is a power of* 3.

(b) For every $d \geq 4$, there is a Steiner triple system with 3^d points which is not isomorphic to $\mathrm{AG}(d,3)$.

Exercises

1. Prove that the number of points in a Steiner triple system is either 0 or congruent to 1 or 3 (mod 6), while the number of points in a Steiner quadruple system is 0, 1, or congruent to 2 or 4 (mod 6).

(It is known that these conditions are sufficient for the existence of Steiner triple and quadruple systems.)

2. Let $(X, \mathcal{L})$ be a Steiner triple system satisfying (AS1). For each point $x \in X$, let τ_x be the permutation of X which fixes x and interchanges y and z whenever $\{x,y,z\}$ is a triple. Prove that

(a) τ_x is an automorphism;

(b) $\tau_x^2 = 1$;

(c) for $x \neq y$, $(\tau_x \tau_y)^3 = 1$.

4

Various topics

This chapter collects some topics, any of which could be expanded into an entire chapter (or even a book!): spreads and translation planes; subsets of projective spaces; projective lines; and the simplicity of $\mathrm{PSL}(n,F)$.

4.1 Spreads and translation planes

Let V be a vector space over F, having even rank $2n$. A *spread* $\mathcal{S}$ is a set of subspaces of V of rank n, having the property that any non-zero vector of V lies in a unique member of $\mathcal{S}$. A trivial example occurs when $n = 1$ and $\mathcal{S}$ consists of all the rank 1 subspaces.

The importance of spreads comes from the following result, whose proof is straightforward.

Proposition 4.1 *Let $\mathcal{S}$ be a spread in V, and $\mathcal{L}$ the set of all cosets of members of $\mathcal{S}$. Then $(V,\mathcal{L})$ is an affine plane. The projective plane obtained by adding a line at infinity L_∞ is (p,L_∞)-transitive for all $p \in L_\infty$.* ■

For finite planes, the converse of the last statement is also true. An affine plane with the property that the projective completion is (p,L_∞)-transitive for all $p \in L_\infty$ is called a *translation plane*.

Example. Let K be an extension field of F with degree n. Take V to be a rank 2 vector space over K, and $\mathcal{S}$ the set of rank 1 K-subspaces. Then, of course, the resulting affine plane is $\mathrm{AG}(2,K)$. Now forget the K-structure, and regard V

as an F- vector space. Such a spread is called *Desarguesian*, because it can be recognised by the fact that the affine plane is Desarguesian.

Projectively, a spread is a set of $(n-1)$-dimensional flats in $\mathrm{PG}(2n-1,F)$, which partitions the points of F. We will examine further the case $n=1$, which will be considered again in section 4.5. **Assume that F is commutative.**

Lemma 4.2 *Given three pairwise skew lines in* $\mathrm{PG}(3,F)$*, there is a unique common transversal through any point on one of the lines.*

Proof Let L_1, L_2, L_3 be the lines, and $p \in L_1$. The quotient space by p is a projective plane $\mathrm{PG}(2,F)$, and $\Pi_1 = \langle p, L_2 \rangle$ and $\Pi_2 = \langle p, L_3 \rangle$ are distinct lines in this plane; they meet in a unique point, which corresponds to a line M containing p and lying in Π_1 and Π_2, hence meeting L_2 and L_3. ■

Now let $\mathcal{R}'$ be the set of common transversals to the three pairwise skew lines. The lines in $\mathcal{R}'$ are pairwise skew, by 4.2.

Lemma 4.3 *A common transversal to three lines of $\mathcal{R}'$ is a transversal to all of them.* ■

For the proof, see Exercise 2, or Section 8.4.

Let $\mathcal{R}$ be the set of all common transversals to $\mathcal{R}'$. The set $\mathcal{R}$ is called a *regulus*, and $\mathcal{R}'$ (which is also a regulus) is the *opposite regulus*. Thus, three pairwise skew lines lie in a unique regulus.

A spread is *regular* if it contains the regulus through any three of its lines.

Theorem 4.4 *A spread is Desarguesian if and only if it is regular.* ■

(The proof of the forward implication is given in Exercise 2.)

If we take a regular spread, and replace the lines in a regulus in this spread by those in the opposite regulus, the result is still a spread; for a regulus and its opposite cover the same set of points. This process is referred to as *derivation*. It gives rise to non-Desarguesian translation planes:

Proposition 4.5 *If $|F| > 2$, then a derivation of a regular spread is not regular.*

Proof Choose two reguli $\mathcal{R}_1$, $\mathcal{R}_2$ with a unique line in common. If we replace $\mathcal{R}_1$ by its opposite, then the regulus $\mathcal{R}_2$ contains three lines of the spread but is not contained in the spread. ■

It is possible to push this much further. For example, any set of pairwise disjoint reguli can be replaced by their opposites. I will not discuss this any further.

The concept of a spread of lines in $\mathrm{PG}(3,F)$ can be dualised. (For the rest of the section, F is not assumed commutative.) A set $\mathcal{S}$ of pairwise skew lines is called a *cospread* if every plane contains a (unique) line of $\mathcal{S}$; in other words, if $\mathcal{S}$ corresponds to a spread in the dual space $\mathrm{PG}(2,F^\circ)$. Call $\mathcal{S}$ a *bispread* if it is both a spread and a cospread.

If F is finite, then every spread is a bispread. (For there are equally many, viz. $(q+1)(q^2+1)$, points and planes; and a set of n pairwise skew lines accounts for $(q+1)n$ points and the same number of planes.) Moreover, a Desarguesian spread is a bispread; and any derivation of a bispread is a bispread (since the concept of a regulus is self-dual). The reader may be wondering if there are any spreads which are not bispreads! That they exist in profusion is a consequence of the next result (take $\mathcal{P}=\emptyset$), and gives us lots of strange translation planes.

Theorem 4.6 *Let F be an infinite field. Let $\mathcal{P}$, $\mathcal{Q}$ be sets of points and planes in $\mathrm{PG}(3,F)$, with the property that $|\mathcal{P}|+|\mathcal{Q}|<|F|$. Then there is a set $\mathcal{S}$ of pairwise skew lines, satisfying*

(a) the point p lies on a line of $\mathcal{S}$ if and only if $p\notin\mathcal{P}$;

(b) the plane Π contains a line of $\mathcal{S}$ if and only if $\Pi\notin\mathcal{Q}$.

Proof We use the fact that $\mathrm{PG}(2,F)$ is not the union of fewer than $|F|$ points and lines. For, if S is any set of fewer than $|F|$ points and lines, and L is a line not in S, then L is not covered by its intersections with members of S.

The proof is a simple transfinite induction. (Note that we are using the Axiom of Choice here; but, in any case, the proof is valid over any field which can be well-ordered, in particular, over any countable field.) For readers unfamiliar with set theory, assume that F is countable, delete the word "transfinite", and ignore comments about limit ordinals in the following argument.

Let α be the initial ordinal of cardinality $|F|$. Well-order the points of $\mathrm{PG}(3,F)$ not in $\mathcal{P}$ and the planes not in $\mathcal{Q}$ in a single sequence of order-type α, say $(X_\beta:\beta<\alpha)$. Construct a sequence $(\mathcal{S}_\beta:\beta<\alpha)$ by transfinite recursion, as follows.

Set $\mathcal{S}_0=\emptyset$.

Suppose that β is a successor ordinal, say $\beta=\gamma+1$. Suppose that X_β is a point (the other case is dual). If $\mathcal{S}_\gamma$ contains a line incident with X_β, then set $\mathcal{S}_\beta=\mathcal{S}_\gamma$. Suppose not. Consider the projective plane $\mathrm{PG}(3,F)/X_\beta$. By our initial remark,

this plane is not covered by fewer than α lines of the form $\langle L, X_\beta\rangle/X_\beta$ (for $L \in \mathcal{S}_\gamma$) or Π/X_β (for $\Pi \in \mathcal{Q}$ with $X_\beta \in \Pi$) and points $\langle p, X_\beta\rangle/X_\beta$ (for $p \in \mathcal{P}$). So we can choose a point lying outside the union of these points and lines, that is, a line L_β containing X_β so that $L_\beta \cap L = \emptyset$ (for $L \in \mathcal{S}_\gamma$), $L_\beta \notin \Pi$ (for $\Pi \in \mathcal{Q}$), and $p \notin L_\beta$ (for $p \in \mathcal{P}$). Set $\mathcal{S}_\beta = \mathcal{S}_\gamma \cup \{L_\beta\}$.

If β is a limit ordinal, set

$$\mathcal{S}_\beta = \bigcup_{\gamma<\beta} \mathcal{S}_\gamma.$$

Then $\mathcal{S}_\alpha$ is the required set of lines. ■

Exercises

1. Show that, if three pairwise skew lines in $\mathrm{PG}(3,F)$ are given, then it is possible to choose coordinates so that the lines have equations

$x_1 = x_2 = 0$;

$x_3 = x_4 = 0$;

$x_3 = x_1$, $x_4 = x_2$.

Find the common transversals to these three lines.

2. Now let F be commutative. Show that the common transversals to any three of the lines found in the last question are the original three lines and the lines with equations

$x_1 = x_3\alpha$, $x_2 = x_4\alpha$

for $\alpha \in F$, $\alpha \neq 0, 1$.

Deduce that the Desarguesian spread defined by a quadratic extension of F is regular.

3. Prove that Lemma 4.3 is valid in $\mathrm{PG}(3,F)$ if and only if F is commutative.

4. Use Theorem 4.6 to show that, if F is an infinite field, then there is a spread of lines in $\mathrm{AG}(3,F)$ which contains one line from each parallel class.

4.2 Some subsets of projective spaces

For most of the second half of these lecture notes, we will be considering subsets of projective spaces which consist of the points (and general subspaces) on

which certain forms vanish identically. In this section, I will describe some more basic subsets of projective spaces, and how to recognise them by their intersections with lines. The first example is a fact we have already met.

Proposition 4.7 *(a) A set S of points in a projective space is a subspace if and only if, for any line L, S contains no point, one point, or all points of L.*

(b) A set S of points in a projective space is a hyperplane if and only if, for any line L, S contains one or all points of L. ■

The main theorem of this section is a generalisation of Proposition 4.7(a). What if we make the condition symmetric, that is, ask that S contains none, one, all but one, or all points of any line L? The result is easiest to state in the finite case:

Theorem 4.8 *Let S be a set of points of $X = \mathrm{PG}(n,F)$ such that, for any line L, S contains none, one, all but one, or all points of S. Suppose that $|F| > 2$. Then there is a chain*

$$\emptyset = X_0 \subset X_1 \subset \ldots \subset X_m = X$$

of subspaces of X, such that either $S = \bigcup_{i\geq 0}(X_{2i+1} \setminus X_{2i})$, or $S = \bigcup_{i\geq 0}(X_{2i+2} \setminus X_{2i+1})$.

The hypothesis that $|F| > 2$ is necessary: over the field $\mathrm{GF}(2)$, a line has just three points, so the four possibilities listed in the hypothesis cover all subsets of a line. This means that any subset of the projective space satisfies the hypothesis! (Nevertheless, see Theorem 4.10 below.)

Note that the hypothesis on S is "self-complementary", and the conclusion must reflect this. It is more natural to talk about a colouring of the points with two colours such that each colour class satisfies the hypothesis of the theorem. In this language, the result can be stated as follows.

Theorem 4.9 *Let the points of a (possibly infinite) projective space X over F be coloured with two colours c_1 and c_2, such that every colour class contains none, one, all but one, or all points of any line. Suppose that $|F| > 2$. Then there is a chain $\mathcal{C}$ of subspaces of X, and a function $f : \mathcal{C} \to \{c_1, c_2\}$, so that*

(a) $\bigcup \mathcal{C} = X$;

(b) for $Y \in \mathcal{C}$, there exist points of Y lying in no smaller subspace in $\mathcal{C}$, and all such points have colour $f(Y)$.

The proof proceeds in a number of stages.

Step 1 The result is true for a projective plane (Exercise 1).

Now we define four relations $<_1$, $<_2$, $<$, $\|$ on X, as follows:

- $p <_1 q$ if p is the only point of its colour on the line pq; (this relation or its converse holds between p and q if and only if p and q have different colours);
- $p <_2 q$ if there exists q with $p <_1 r$ and $r <_1 q$ (this holds only if p and q have the same colour);
- $p < q$ if $p <_1 q$ or $p <_2 q$;
- $p \| q$ if neither $p < q$ nor $q < p$ (this holds only if p and q have the same colour).

Step 2 There do not exist points p, q with $p <_2 q$ and $q <_2 p$.

For, if so, then (with $p_1 = p$, $q_1 = q$) there are points p_2, q_2 such that

$$p_1 <_1 p_2 <_1 q_1 <_1 q_2 <_2 p_1.$$

Let c_i be the colour of p_i and q_i, $i = 1, 2$. By Step 1, the colouring of the plane $p_1p_2q_1$ is determined; and every point of this plane off the line p_1p_2. In particular, if $x_1 \in p_1q_1$, $x_1 \neq p_1, q_1$, then every point of x_1p_2 except p_2 has colour c_1. Similarly, every point of x_1q_2 except q_2 has colour c_1; and then every point of x_1x_2 except x_2 has colour c_1, where $x_2 \in p_2q_2$, $x_2 \neq p_2, q_2$.

But, by the same argument, every point of x_1x_2 except x_1 has colour c_2, giving a contradiction.

Step 3 $<$ is a partial order.

The antisymmetry follows by definition for $<_1$ and by Step 2 for $<_2$; we must prove transitivity. So suppose that $p < q < r$, and consider cases. If $p <_1 q <_1 r$, then $p <_2 r$ by definition. If $p <_1 q <_2 r$ or $p <_2 q <_1 r$, then p and r have different colours and so are comparable; and $r <_1 p$ contradicts Step 2. Finally, if $p <_2 q <_2 r$, then $p <_1 s <_1 q$ for some s; then $s <_1 r$, so that $p <_2 r$.

Step 4 If $p < q \| r$ or $p \| q < r$, then $p < r$; and if $p \| q \| r$, then $p \| r$.

Suppose that $p < q \| r$. If $p <_1 q$, then p and r have different colours and so are comparable; and $r <_1 p$ would imply $r < q$ by Step 3, so $p <_1 r$. The next case is similar. The last assertion is a simple consequence of the other two.

Step 5 If $p < q$, then $p < r$ for all points r of pq except p; and the points of pq other than p are pairwise incomparable.

This holds by assumption if $p <_1 q$, and by the proof of Step 2 if $p <_2 q$.

Now let $S(p) = \{q : p \not< q\}$, and $T(p) = \{q : q < p\}$.

Step 6 $S(p)$ and $T(p)$ are subspaces, with $p \in S(p) \setminus T(p)$. Moreover, $T(p)$ is the union of the spaces $S(q)$ for $q < p$, and is spanned by the points of $S(p)$ with colour different from that of p; and we have

$p \| q$ implies $S(p) = S(q)$;

$q < p$ implies $S(q) \subseteq T(p)$.

All of this follows by straightforward argument from the preceding steps.

Now the proof of the Theorem follows: we set $\mathcal{C} = \{S(p) : p \in X\}$, and let $f(S(p))$ be the colour of p. The conclusions of the Theorem follow from the assertions in Step 6. ■

Remark. The only place in the above argument where the hypothesis $|F| > 2$ was used was in Step 1. Now $\mathrm{PG}(2,2)$ has seven points; so, up to complementation, a subset of $\mathrm{PG}(2,2)$ is empty, a point, a line with a point removed, a line, or a triangle. Only the last case fails to satisfy the conclusion of the Theorem. So we have the following result:

Theorem 4.10 *The conclusions of Theorems 4.8 and 4.9 remain true in the case $F = \mathrm{GF}(2)$ provided that we add the extra hypothesis that no colour class intersects a plane in a triangle (or, in 4.8, that no plane meets S in a triangle or the complement of one).* ■

Exercise

1. Prove that Theorem 4.8 holds in any projective plane of order greater than 2 (not necessarily Desarguesian).

4.3 Segre's Theorem

For projective geometries over finite fields, it is very natural to ask for characterisations of interesting sets of points by hypotheses on their intersections with

lines. Very much finer discriminations are possible with finite than with infinite cardinal numbers; for example, all infinite subsets of a countably infinite set whose complements are also infinite are alike.

It is not my intention to survey even a small part of this vast literature. But I will describe one of the earliest and most celebrated results of this kind. I begin with some generalities about algebraic curves. Assume that F is a *commutative* field.

If a polynomial f in $x_1,\ldots,x_{n+1}$ is *homogeneous*, that is, a sum of terms all of the same degree, then $f(\mathbf{v})=0$ implies $f(\alpha\mathbf{v})=0$ for all $\alpha\in F$. So, it f vanishes at a non-zero vector, then it vanishes at the rank 1 subspace (the point of $\mathrm{PG}(n,F)$) it spans. The *algebraic variety* defined by f is the set of points spanned by zeros of f. We are concerned here only with the case $n=2$, in which case (assuming that f does not vanish identically) this set is called an *algebraic curve*.

Now consider the case where f has degree 2, and $F=\mathrm{GF}(q)$, where q is an *odd* prime power. The curve it defines may be a single point, or a line, or two lines; but, if none of these occurs, then it is equivalent (under the group $\mathrm{PGL}(3,q)$) to the curve defined by the equation $x_1^2+x_2^2+x_3^2=0$ (see Exercise 1). Any curve equivalent to this one is called a *conic* (or *irreducible conic*).

It can be shown (see Exercise 2) that a conic has $q+1$ points, no three of which are collinear. The converse assertion is the content of Segre's Theorem:

Theorem 4.11 (Segre's Theorem) *For q odd, a set of $q+1$ points in $\mathrm{PG}(2,q)$, with no three collinear, is a conic.*

Proof Let $\mathcal{O}$ be an oval. We begin with some combinatorial analysis which applies in any plane of odd order; then we introduce coordinates.

Step 1 Any point not on $\mathcal{O}$ lies on 0 or 2 tangents.

Proof Let p be a point not on $\mathcal{O}$. Since $|\mathcal{O}|=q+1$ is even, and an even number of points lie on secants through p, an even number must lie on tangents also. Let x_i be the number of points outside $\mathcal{O}$ which lie on i tangents. Now we have

$$\begin{aligned}\sum x_i &= q^2,\\ \sum ix_i &= (q+1)q,\\ \sum i(i-1)x_i &= (q+1)q.\end{aligned}$$

(These are all obtained by double counting. The first holds because there are q^2 points outside $\mathcal{O}$; the second because there are $q+1$ tangents (one at each point

of O), each containing q points not on O; and the third because any two tangents intersect at a unique point outside O.)

From these equations, we see that $\sum i(i-2)x_i = 0$. But the term $i = 1$ in the sum vanishes (any point lies on an even number of tangents); the terms $i = 0$ and $i = 2$ clearly vanish, and $i(i-2) > 0$ for any other value of i. So $x_i = 0$ for all $i \neq 0$ or 2, proving the assertion.

Remark Points not on O are called *exterior points* or *interior points* according as they lie on 2 or 0 tangents, by analogy with the real case. But the analogy goes no further. In the real case, every line through an interior point is a secant; this is false for finite planes.

Step 2 The product of all the non-zero elements of $\mathrm{GF}(q)$ is equal to -1.

Proof The solutions of the quadratic $x^2 = 1$ are $x = 1$ and $x = -1$; these are the only elements equal to their multiplicative inverses. So, in the product of all the non-zero elements, everything except 1 and -1 pairs off with its inverse, leaving these two elements unpaired.

For the next two steps, note that we can choose the coordinate system so that the sides of a given triangle have equations $x = 0$, $y = 0$ and $z = 0$ (and the opposite vertices are $[1,0,0]$, $[0,1,0]$, and $[0,0,1]$ respectively). We'll call this the *triangle of reference.*

Step 3 Suppose that concurrent lines through the vertices of the triangle of reference meet the opposite sides in the points $[0,1,a]$, $[b,0,1]$, and $[1,c,0]$. Then $abc = 1$.

Proof The equations of the concurrent lines are $z = ay$, $x = bz$ and $y = cx$ respectively; the point of concurrency must satisfy all three equations, whence $abc = 1$.)

Remark This result is equivalent to the classical Theorem of Menelaus.

Step 4 Let the vertices of the triangle of reference be chosen to be three points of O, and let the tangents at these points have equations $z = ay$, $x = bz$ and $y = cx$ respectively. Then $abc = -1$.

Proof There are $q-2$ further points of $\mathcal{O}$, say $p_1,\dots,p_{q-2}$. Consider the point $[1,0,0]$. It lies on the tangent $z=ay$, meeting the opposite side in $[0,1,a]$; two secants which are sides of the triangle; and $q-2$ further secants, through $p_1,\dots,p_{q-2}$. Let the secant through p_i meet the opposite side in $[0,1,a_i]$. Then $a\prod_{i=1}^{q-2}a_i=-1$, by Step 2. If b_i,c_i are similarly defined, we have also $b\prod_{i=1}^{q-2}b_i=c\prod_{i=1}^{q-2}c_i=-1$. Thus

$$abc\prod_{i=1}^{q-2}(a_ib_ic_i)=-1.$$

But, by Step 3, $a_ib_ic_i=1$ for $i=1,\dots,q-2$; so $abc=-1$.

Step 5 Given any three points p,q,r of $\mathcal{O}$, there is a conic $\mathcal{C}$ passing through p,q,r and having the same tangents at these points as does $\mathcal{O}$.

Proof Choosing coordinates as in Step 4, the conic with equation

$$yz-czx+caxy=0$$

can be checked to have the required property. (For example, $[1,0,0]$ lies on this conic; and, putting $z=ay$, we obtain $ay^2=0$, so $[1,0,0]$ is the unique point of the conic on this line.)

Step 6 Now we are finished if we can show that the conic $\mathcal{C}$ of Step 5 passes through an arbitrary further point s of $\mathcal{O}$.

Proof Let $\mathcal{C}'$ and $\mathcal{C}''$ be the conics passing through p,q,s and p,r,s respectively and having the correct tangents there. Let the conics $\mathcal{C}$, $\mathcal{C}'$ and $\mathcal{C}''$ have equations $f=0$, $f'=0$, $f''=0$ respectively. (These equations are determined up to a constant factor.) Let L_p,L_q,L_r,L_s be the tangents to $\mathcal{O}$ at p,q,r,s respectively. Since all three conics are tangent to L_p at p, we can choose the normalisation so that f,f',f'' agree identically on L_p.

Now consider the restrictions of f' and f'' to L_s. Both are quadratic functions having a double zero at s, and the values at the point $L_s\cap L_p$ coincide; so the two functions agree identically on L_s. Similarly, f and f' agree on L_q, and f and f'' agree on L_r. But then f, f' and f'' all agree at the point $L_q\cap L_r$. So the quadratic functions f' and f'' agree on L_p, L_s, and $L_q\cap L_r$, which forces them to be equal. So the three conics coincide, and our claim is proved (and with it Segre's Theorem). ∎

The argument in the last part of the proof can be generalised to give the following result (of which it forms the case $n = q+1$, $m = 2$, with $L_1, \ldots, L_n$ the tangents to the oval, and $\{p_{i1}, p_{i2}\}$ the point of tangency of L_i taken with multiplicity 2).

Proposition 4.12 *Let $L_1, \ldots, L_n$ be lines in $\mathrm{PG}(2,q)$, no three concurrent. Let $p_{i1}, \ldots, p_{im}$ be points of L_i, not necessarily distinct, but lying on none of the other $L_{i'}$. Suppose that, for any three of the lines, there is an algebraic curve of degree m whose intersections with those lines are precisely the specified points (counted with the appropriate multiplicity). Then there is a curve of degree m, meeting each line in just the specified points.* ■

Proposition 4.12 has been generalised [32] to arbitrary sets of lines (without the assumption that no three are concurrent).

Proposition 4.13 *Let $L_1, \ldots, L_n$ be lines in $\mathrm{PG}(2,q)$. Let $p_{i1}, \ldots, p_{im}$ be points of L_i, not necessarily distinct, but lying on none of the other $L_{i'}$. Suppose that, for any three of the lines which form a triangle, and for the set of all lines passing through any point of the plane (whenever there are at least three such lines), there is an algebraic curve of degree m whose intersections with those lines are precisely the specified points (counted with the appropriate multiplicity). Then there is a curve of degree m, meeting each line in just the specified points.* ■

The analogue of Segre's Theorem over $\mathrm{GF}(q)$ with even q is false. In this case, the tangents to an oval S all pass through a single point n, the *nucleus* of the oval (Exercise 4); and, for any $p \in S$, the set $S \cup \{n\} \setminus \{p\}$ is also an oval. But, if $q > 4$, then at most one of these ovals can be a conic (see Exercise 5: these ovals have q common points). For sufficiently large q (viz., $q \geq 64$), and also for $q = 16$, there are other ovals, not arising from this construction. We refer to [3] or [14] for up-to-date information on ovals in planes of even order.

We saw that there are ovals in infinite projective planes which are not conics. However, there is a remarkable characterisation of conics due to Buekenhout. A hexagon is said to be *Pascalian* if the three points of intersection of opposite sides are collinear. In this terminology, Pappus' Theorem asserts that a hexagon whose vertices lie alternately on two lines is Pascalian. Since a pair of lines forms a "degenerate conic", this theorem is generalised by Pascal's Theorem:

Theorem 4.14 (Pascal's Theorem) *In a Pappian projective plane, a hexagon inscribed in a conic is Pascalian.* ■

We know from Theorem 2.3 that a projective plane satisfying Pappus' Theorem is isomorphic to $\mathrm{PG}(2,F)$ for a commutative field F. The theorem of Buekenhout completes this circle of ideas. Its proof is group-theoretic, using a characterisation of $\mathrm{PGL}(2,F)$ as sharply 3-transitive group due to Tits.

Theorem 4.15 (Buekenhout's Theorem) *Let S be an oval in a projective plane Π. Suppose that every hexagon with vertices in S is Pascalian. Then Π is isomorphic to $\mathrm{PG}(2,F)$ for some commutative field F, and S is a conic in Π.* ■

Exercises

1. (a) By completing the square, prove that any homogeneous polynomial of degree 2 in n variables, over a commutative field F with characteristic different from 2, is equivalent (by non-singular linear transformation) to the polynomial

$$\alpha_1 x_1^2 + \ldots + \alpha_n x_n^2.$$

(b) Prove that multiplication of any α_i in the above form by a square in F gives an equivalent form.

(c) Now let $F = \mathrm{GF}(q)$ and $n = 3$; let η be a fixed nonsquare in F. Show that the curves defined by x_1^2, $x_1^2 - x_2^2$ and $x_1^2 - \eta x_2^2$ are respectively a line, two lines, and a point. Show that there exists α such that $\eta = 1 + \alpha^2$. Observing that

$$(x+\alpha y)^2 + (\alpha x - y)^2 = \eta(x^2+y^2),$$

prove that the forms $x_1^2 + x_2^2 + x_3^2$ and $x_1^2 + \eta x_2^2 + \eta x_3^2$ are equivalent. Deduce the classification of curves of degree 2 over $\mathrm{GF}(q)$ given in the text.

2. Count the number of secants through an exterior point and through an interior point of an oval in a projective plane of odd order q. Also, count the number of points of each type.

3. Prove that a curve of degree 2 over any commutative field is empty, a point, a line, a pair of lines, or an oval. Prove also that a curve of degree 2 over a finite field is non-empty.

4. Prove that, if q is even, then the tangents to an oval in a projective plane of order q are concurrent. Deduce that there is a set of $q+2$ points with no three collinear, having no tangents (i.e., meeting every line in 0 or 2 points). (Removing any one of these points then gives an oval.)

Remark. A set of $n+2$ points in a plane of order n, no three collinear, is called a *hyperoval*.

5. Prove that, in any infinite projective plane, for any integer $k > 1$, there is a set of points meeting every line in exactly k points.

6. Prove that five points of $\mathrm{PG}(2,F)$, with no three collinear, are contained in a unique conic. (Take four of the points to be the standard set $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ and $(1,1,1)$; the fifth is $(1,\alpha,\beta)$, where α and β are distinct from one another and from 0 and 1.)

4.4 Ovoids and inversive planes

Ovoids are 3-dimensional analogues of ovals. They have added importance because of their connection with inversive planes, which are one-point extensions of affine planes. (The traditional example is the relation between the Riemann sphere and the "extended complex plane".)

Fields in this section are commutative.

An *ovoid* in $\mathrm{PG}(3,F)$ is a set O of points with the properties

(O1) no three points of O are collinear;

(O2) the tangents to O through a point of O form a plane pencil.

(If a set of points satisfies (O1), a line is called a *secant*, *tangent* or *passant* if it meets the set in 2, 1 or 0 points respectively. The plane containing the tangents to an ovoid at a point x is called the *tangent plane* at x.)

The classical examples of ovoids are the *elliptic quadrics*. Let $\alpha x^2+\beta x+\gamma$ be an irreducible quadratic over the field F. The elliptic quadric consists of the points of $\mathrm{PG}(3,F)$ whose coordinates (x_1,x_2,x_3,x_4) satisfy

$$x_1x_2+\alpha x_3^2+\beta x_3x_4+\gamma x_4^2=0.$$

The proof that these points do form an ovoid is left as an exercise.

Over finite fields, ovoids are rare. Barlotti and Panella showed the following analogue of Segre's theorem on ovals:

Theorem 4.16 *Any ovoid in* $\mathrm{PG}(3,q)$*, for q an odd prime power, is an elliptic quadric.* ■

For even q, just one further family is known, the *Suzuki–Tits ovoids*, which we will construct in Section 8.4.

An *inversive plane* is, as said above, a one-point extension of an affine plane. That is, it is a pair $(X, \mathcal{C})$, where X is a set of points, and $\mathcal{C}$ a collection of subsets of X called *circles*, satisfying

(I1) any three points lie in a unique circle;

(I2) if x, y are points and C a circle with $x \in C$ and $y \notin C$, then there is a unique circle C' satisfying $y \in C'$ and $C \cap C' = \{x\}$;

(I3) there exist four non-concircular points.

It is readily checked that, for $x \in X$, the points different from x and circles containing x form an affine plane. The *order* of the inversive plane is the (common) order of its derived affine planes.

Proposition 4.17 *The points and non-trivial plane sections of an ovoid form an inversive plane.*

Proof A plane section of the ovoid O is non-trivial if it contains more than one point. Any three points of O are non-collinear, and so define a unique plane section. Given x, the points of O different from x and the circles containing x correspond to the lines through x not in the tangent plane T_x and the planes through x different from T_x; these are the points of the quotient space not incident with the line T_x/x and the lines different from T_x/x, which form an affine plane. ■

An inversive plane arising from an ovoid in this way is called *egglike*. Dembowski proved:

Theorem 4.18 *Any inversive plane of even order is egglike (and so its order is a power of 2).* ■

This is not known to hold for odd order, but no counterexamples are known.

There are configuration theorems (the *bundle theorem* and *Miquel's theorem* respectively) which characterise egglike inversive planes and "classical" inversive planes (coming from the elliptic quadric) respectively.

Higher-dimensional objects can also be defined. A set O of points of $\mathrm{PG}(n, F)$ is an *ovoid* if

(O1) no three points of O are collinear;

(O2') the tangents to $\mathcal{O}$ through a point x of $\mathcal{O}$ are all the lines through x in a hyperplane of $\mathrm{PG}(n,F)$.

Proposition 4.19 *If F is finite and $n \geq 4$, then $\mathrm{PG}(n,F)$ contains no ovoid.* ∎

However, there can exist such ovoids over infinite fields (Exercise 3).

Exercises

1. Prove Proposition 4.19. [Hint: it suffices to prove it for $n = 4$.]

2. Prove that, for q odd, a set of points in $\mathrm{PG}(3,q)$ which satisfies (O1) has cardinality at most q^2+1, with equality if and only if it is an ovoid.

(This is true for q even, $q > 2$ also, though the proof is much harder. For $q = 2$, the complement of a hyperplane is a set of 8 points in $\mathrm{PG}(3,2)$ satisfying (O1).)

3. Show that the set of points of $\mathrm{PG}(n,\mathbb{R})$ whose coordinates satisfy

$$x_1x_2 + x_3^2 + \ldots + x_n^2 = 0$$

is an ovoid.

4.5 Projective lines

A projective line over a field F has no non-trivial structure as an incidence geometry. From the Kleinian point of view, though, it does have geometric structure, derived from the fact that the group $\mathrm{PGL}(2,F)$ operates on it. As we saw earlier, the action of this group is 3-transitive (sharply so if F is commutative), and can even be 4-transitive for special skew fields of characteristic 2. However, we assume in this section that the field is commutative.

It is conventional to label the points of the projective line over F with elements of $F \cup \{\infty\}$, as follows: the point $\langle(1,\alpha)\rangle$ is labelled by α, and the point $\langle(0,1)\rangle$ by ∞. (If we regard points of $\mathrm{PG}(2,F)$ as lines in the affine plane $\mathrm{AG}(2,F)$, then the label of a point is the slope of the corresponding line.)

Since $\mathrm{PGL}(2,F)$ is sharply 3-transitive, distinguishing three points must give unique descriptions to all the others. This is conveniently done by means of the *cross ratio*, the function from 4-tuples of distinct points to $F \setminus \{0,1\}$, defined by

$$f(x_1,x_2,x_3,x_4) = \frac{(x_1-x_3)(x_4-x_2)}{(x_1-x_4)(x_3-x_2)}.$$

In calculating cross ratio, we use the same conventions for dealing with ∞ as when elements of $\mathrm{PGL}(2,F)$ are represented by linear fractional transformations; for example, $\infty - \alpha = \infty$, and $\alpha\infty/\beta\infty = \alpha/\beta$. Slightly differing forms of the cross ratio are often used; the one given here has the property that $f(\infty,0,1,\alpha) = \alpha$.

Proposition 4.20 *The group of permutations of* $\mathrm{PG}(1,F)$ *preserving the cross ratio is* $\mathrm{PGL}(2,F)$.

Proof Calculation establishes that linear fractional transformations do preserve cross ratio. Also, the cross ratio as a function of its fourth argument, with the first three fixed, is one-to-one, so a permutation which preserves cross ratio and fixes three points is the identity. The result follows from these two assertions. ∎

The cross ratio of four points is unaltered if the arguments are permuted in two cycles of length 2: for example, $f(x_3,x_4,x_1,x_2) = f(x_1,x_2,x_3,x_4)$. These permutations, together with the identity, form a normal subgroup of index six in the symmetric group S_4. Thus, in general, six different values are obtained by permuting the arguments. If α is one of these values, the others are $1-\alpha$, $1/\alpha$, $(\alpha-1)/\alpha$, $1/(1-\alpha)$, and $\alpha/(\alpha-1)$. There are two special cases where the number of values is smaller, that is, where two of the six coincide. The relevant sets are $\{-1,2,\frac{1}{2}\}$, and $\{-\omega,-\omega^2\}$, where ω is a primitive cube root of unity. A quadruple of points is called *harmonic* if its cross ratios belong to the first set, *equianharmonic* if they belong to the second. The first type occurs over any field of characteristic different from 2, while the second occurs only if F contains primitive cube roots of 1. (But note that, if F has characteristic 3, then the two types effectively coincide: $-1 = 2 = \frac{1}{2}$, and the cross ratio of a harmonic quadruple is invariant under all permutations of its arguments!)

In the arguments below, we regard a "quadruple" as being an equivalence class of ordered quadruples (all having the same cross-ratio). So, for example, a harmonic quadruple (in characteristic different from 3) is a 4-set with a distinguished partition into two 2-sets.

Proposition 4.21 *Suppose that the characteristic of* F *is not equal to* 2. *Then the group of permutations which preserve the set of harmonic quadruples is* $\mathrm{P\Gamma L}(2,F)$.

Proof Again, any element of $\mathrm{P\Gamma L}(2,F)$ preserves the set of harmonic quadruples. To see the converse, note that $\mathrm{PGL}(2,F)$ contains a unique conjugacy class

of involutions having two fixed points, and that, if x_1, x_2 are fixed points and (x_3, x_4) a 2-cycle of such an involution, then $\{x_1, x_2, x_3, x_4\}$ is harmonic (and the disthinguished partition is $\{\{x_1, x_2\}, \{x_3, x_4\}\}$). Thus, these involutions can be reconstructed from the set of harmonic quadruples. So any permutation preserving the harmonic quadruples normalises the group G generated by these involutions. We see below that G is $\mathrm{PSL}(2, F)$ if F contains square roots of -1, or contains this group as a subgroup of index 2 otherwise. The normaliser of G is thus $\mathrm{P\Gamma L}(2, F)$, as required. ■

($\mathrm{PSL}(n, F)$ is the group induced on the projective space by the invertible linear transformations with determinant 1.)

We look further at the claim about G in the above proof. A *transvection* is a linear transformation g with all eigenvalues equal to 1, for which $\ker(g-1)$ has codimension 1. In our present case, any 2×2 upper unitriangular matrix different from the identity is a transvection. The collineation of projective space induced by a transvection is called an *elation*. An elation is characterised by the fact that its fixed points form a hyperplane, known as the *axis* of the elation. Dually, an elation fixes every line through a point, called the *centre* of the elation, which is incident with the axis. In the present case $n = 2$, the centre and axis of an elation coincide.

Proposition 4.22 *The elations in* $\mathrm{PGL}(2, F)$ *generate* $\mathrm{PSL}(2, F)$.

Proof The elations fixing a specified point, together with the identity, form a group which acts sharply transitively on the remaining points. Hence the group generated by the elations is 2-transitive. If $\alpha = -1 - 1/\beta$ and $\gamma = \alpha/\beta$, then

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix} = \begin{pmatrix} -1/\beta & 0 \\ 0 & -\beta \end{pmatrix},$$

so the two-point stabiliser in the group generated by all the elations contains that in $\mathrm{PSL}(2, F)$. But elations have determinant 1, and so the group they generate is a subgroup of $\mathrm{PSL}(2, F)$. So we have equality. ■

Now, if two distinct involutions have a common fixed point, then their product is a elation. Since all elations are conjugate, all can be realised in this way. Thus the group G in the proof of Proposition 4.21 contains all elations, and hence contains $\mathrm{PSL}(2, F)$.

We conclude with a different way of giving structure to the projective line. Suppose that E is a subfield of F. Then $\{\infty\} \cup E$ is a subset of the projective line $\{\infty\} \cup F$ having the structure (in any of the senses previously defined) of projective line over E. We call any image of this set under an element of $\mathrm{PGL}(2,F)$ a *circle*. Then any three points lie in a unique circle. The points and circles form an incidence structure which is an extension of the point-line structure of affine space $\mathrm{AG}(n,E)$, where n is the degree of F over E. (For consider the blocks containing ∞. On removing the point ∞, we can regard F as an E-vector space of rank n; E itself is an affine line, and the elements of $\mathrm{PGL}(2,F)$ fixing ∞ are affine transformations; so, for any circle C containing ∞, $C \setminus \{\infty\}$ is an affine line. Since three points lie in a unique circle, every affine line arises in this way.)

Sometimes, as we will see, this geometry can be represented as the points and plane sections of a quadric over E. the most familiar example is the Riemann sphere, which is the projective line over $\mathbb{C}$, and can be identified with a sphere in real 3-space so that the "circles" are plane sections.

4.6 Generation and simplicity

In this section, we extend to arbitrary rank the statement that $\mathrm{PSL}(n,F)$ is generated by elations, and show that this group is simple, except in two special cases.

As before, F is a commutative field.

Theorem 4.23 *For any $n \geq 2$, the group $\mathrm{PSL}(n,F)$ is generated by all elations.*

Proof We use induction on n, the case $n = 2$ having been settled by Proposition 4.22. The induction is based on the fact that, if W is a subspace of the axis of an elation g, then g induces an elation on the quotient projective space modulo W. Given $g \in \mathrm{PSL}(n,F)$, with $g \neq 1$, we have to express g as a product of elations. We may suppose that g fixes a point x. (For, if $xg = y \neq x$, and h is any elation mapping x to y, then gh^{-1} fixes x, and gh^{-1} is a product of elations if and only if g is.

By induction, we may multiply g by a product of elations (whose axes contain x) to obtain an element fixing every line through x; so we may assume that g itself does so. Considering a matrix representing g, and using the fact that $g \in \mathrm{PSL}(n,F)$, we see that g is an elation. ∎

Theorem 4.24 *Suppose that either $n \geq 3$, or $n = 2$ and $|F| > 3$. Then any non-trivial normal subgroup of $\mathrm{PGL}(n,F)$ contains $\mathrm{PSL}(n,F)$.*

Proof We begin with an observation — if N is a normal subgroup of G, and $g \in N$, $g_1 \in G$, then $[g,g_1] \in N$, where $[g,g_1] = g^{-1}g_1^{-1}gg - 1$ is the *commutator* of g and g_1 — and a lemma:

Lemma 4.25 *Under the hypotheses of Theorem 4.24, if $g \in \mathrm{PGL}(n,F)$ maps the point p_1 of $\mathrm{PG}(n-1,F)$ to the point p_2, then there exists $g_1 \in PGL(n,F)$ which fixes p_1 and p_2 and doesn't commute with g.*

Proof *Case 1*: $p_2 g = p_3 \neq p_2$. We can choose g_1 to fix p_1 and p_2 and move p_3. (If p_1, p_2, p_3 are not collinear, this is clear. If they are collinear, use the fact that $\mathrm{PGL}(2,F)$ is 3-transitive on the projective line, which has more than three points.

Case 2: $p_2 g = p_1$. Then g fixes the line $p_1 p_2$, and we can choose coordinates on this line so that $p_1 = \infty$, $p_2 = 0$. Now g acts as $x \mapsto \alpha/x$ for some $\alpha \in F$. Let g_1 induce $x \mapsto \beta x$ on this line; then $[g,g_1]$ induces $x \mapsto \beta^2 x$. So choose $\beta \neq 0, 1, -1$, as we may since $|F| > 3$. ■

So let N be a non-trivial subgroup of $\mathrm{PGL}(n,F)$. Suppose that $g \in N$ maps the hyperplane H_1 to $H_2 \neq H_1$. By the dual form of the Lemma, there exists g_1 fixing H_1 and H_2 and not commuting with g; then $[g,g_1]$ fixes H_2. So we may assume that $g \in N$ fixes a hyperplane H.

Next, suppose that g doesn't fix H pointwise. The group of elations with axis H is isomorphic to the additive group of a vector space whose associated projective space is H; so there is a transvection g_1 with axis H not commuting with g. Then $[g,g_1]$ fixes H pointwise. So we may assume that g fixes H pointwise.

If g is not an elation, then it is a *homology* (induced by a diagonalisable linear map with two eigenvalues, one having multiplicity $n-1$; equivalently, its fixed points form a hyperplane and one additional point). Now if g_1 is an elation with axis H, then $[g,g_1]$ is a non-identity elation.

We conclude that N contains an elation. But then N contains all elations (since they are conjugate), whence N contains $\mathrm{PSL}(n,F)$. ■

For small n and small finite fields $F = \mathrm{GF}(q)$, the group $\mathrm{PSL}(n,q) = \mathrm{PSL}(n,F)$ is familiar in other guises. For $n = 2$, recall that it is sharply 3-transitive of degree $q+1$. Hence we have $\mathrm{PSL}(2,2) \cong S_3$, $\mathrm{PSL}(2,3) \cong A_4$, and $\mathrm{PSL}(2,4) \cong A_5$ (the alternating groups of degrees 4 and 5 — the former is not simple, the latter is the unique simple group of order 60). Less obviously, $\mathrm{PSL}(2,5) \cong A_5$, since it is also simple of order 60. Furthermore, $\mathrm{PSL}(2,7) \cong \mathrm{PSL}(3,2)$ (the unique simple group of order 168), $\mathrm{PSL}(2,9) \cong A_6$, and $\mathrm{PSL}(4,2) \cong A_8$ (for reasons we will see later).

There has been a lot of work, much of it with a very geometric flavour, concerning groups generated by subsets of the set of elations. For example, McLaughlin [22, 23] found all irreducible groups generated by "full elation subgroups" (all elations with given centre and axis). This result was put in a wider context by Cameron and Hall [11]. (In particular, they extended the result to spaces of infinite dimension.) Note that an important ingredient in the arguments of Cameron and Hall is Theorem 4.9: under slight additional hypotheses, the set of all elation centres satisfies the conditions on a colour class in that theorem. The result of Theorem 4.9, together with the irreducibility of the group, then implies that every point is an elation centre.

Exercises

1. (a) Prove that the non-negative integer m is the number of fixed points of an element of $\mathrm{PGL}(n,q)$ if and only if, when written in the base q, its digits are non-decreasing and have sum not exceeding n.

(b) (Harder) Prove that the non-negative integer m is the number of fixed points of an element of $\mathrm{P\Gamma L}(n,F)$ if and only if there exists r such that q is a power of r and, when m is written in the base r, its digits are non-decreasing and have sum at most n.

2. Prove that a simple group of order 60 possesses five Sylow 2-subgroups, which it permutes by conjugation; deduce that such a group is isomorphic to A_5.

3. Modify the proof of Theorem 4.6.2 to show that, under the same hypotheses, $\mathrm{PSL}(n,F)$ is simple. [It is only necessary to show that the various g_1s can be chosen to lie in $\mathrm{PSL}(n,F)$. The only case where this fails is Case 2 of the Lemma when $n=2$, $F=\mathrm{GF}(5)$.]

4. (a) Let Π be a projective plane of order 4 containing a hyperoval X (six points, no three collinear). Prove that there are natural bijections between the set of lines meeting X in two points and the set of 2-subsets of X; and between the set of points outside X and the set of partitions of X into three 2-subsets. Find a similar description of a set bijective with the set of lines disjoint from X. Hence show that Π is unique (up to isomorphism).

(b) Let Π be a projective plane of order 4. Prove that any four points, no three collinear, are contained in a hyperoval. Hence show that there is a unique projective plane of order 4 (up to isomorphism).

(See Cameron and Van Lint [F] for more on the underlying combinatorial principle.)

5

Buekenhout geometries

Francis Buekenhout introduced an approach to geometry which has the advantages of being both general, and local (a geometry is studied *via* its residues of small rank). In this chapter, we introduce Buekenhout's geometries, and illustrate with projective spaces and related objects. Further examples will occur later (polar spaces).

5.1 Buekenhout geometries

So far, nothing has been said in general about what a "geometry" is. Projective and affine geometries have been defined as collections of subspaces, but even the structure carried by the set of subspaces was left a bit vague (except in Section 3.4, where we used the inclusion partial order to characterise generalised projective spaces as lattices). In this section, I will follow an approach due to Buekenhout (inspired by the early work of Tits on buildings).

Before giving the formal definition, let us remark that the subspaces or flats of a projective geometry are of various types (i.e., of various dimensions); may or may not be incident (two subspaces are incident if one contains the other); and are partially ordered by inclusion. To allow for duality, we do not want to take the partial order as basic; and, as we will see, the betweenness relation derived from it can be deduced from the type and incidence relations. So we regard type and incidence as basic.

A *geometry*, or *Buekenhout geometry*, then, has the following ingredients: a set X of *varieties*, a symmetric *incidence relation* I on X, a finite set Δ of *types*, and a *type map* $\tau : X \to \Delta$. We require the following axiom:

(B1) Two varieties of the same type are incident if and only if they are equal.

In other words, a geometry is a multipartite graph, where we have names for the multipartite blocks ("types") of the graph. We mostly use familiar geometric language for incidence; but sometimes, graph-theoretic terms like diameter and girth will be useful. But one graph-theoretic concept is vital; a geometry is *connected* if the graph of varieties and incidence is connected.

The *rank* of a geometry is the number of types.

A *flag* is a set of pairwise incident varieties. It follows from (B1) that the members of a flag have different types. A geometry satisfies the *transversality condition* if the following strengthening of (B1) holds:

(B2) (a) Every flag is contained in a maximal flag.

(b) Every maximal flag contains one variety of each type.

All geometries here will satisfy transversality.

Let F be a flag in a geometry G. The *residue* $G_F = R(F)$ of F is defined as follows: the set of varieties is

$$X_F = \{x \in X \setminus F : xIy \text{ for all } y \in F\};$$

the set of types is $\Delta_F = \Delta \setminus \tau(F)$; and incidence and the type map are the restrictions of those in G. It satisfies (B1) (resp. (B2)) if G does. The *type* of a flag or residue is its image under the type map, and the *cotype* is the complement of the type in Δ; so the type of G_F is the cotype of F. The *rank* and *corank* are the cardinalities of the type and cotype.

A transversal geometry is called *thick* (resp. *firm thin*) if every flag of corank 1 is contained in at least three (resp. at least two, exactly two) maximal flags.

A property holds *residually* in a geometry if it holds in every residue of rank at least 2. (Residues of rank 1 are sets without structure.) In particular, all geometries of interest are *residually connected*; in effect, we assume residual connectedness as an axiom:

(B3) All residues of rank at least 2 are connected.

The next result illustrates this concept.

Proposition 5.1 *Let G be a residually connected transversal geometry, and let x and y be varieties of X, and i and j distinct types. Then there is a path from x to y in which all varieties except possibly x and y have type i or j.*

Proof The proof is by induction on the rank. For rank 2, residual connectedness is just connectedness, and the result holds by definition. So assume the result for all geometries of smaller rank than G.

We show first that a two-step path whose middle vertex is not of type i or j can be replaced by a path of the type required. So let xzy be a path of length 2. Then x and y lie in the residue of z; so the assertion follows from the inductive hypothesis.

Now this construction reduces by one the number of interior vertices not of type i or j on a path with specified endpoints. Repeating it as often as necessary gives the result. ■

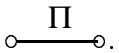
The heart of Buekenhout's idea is that "local" conditions on (or axiomatisations of) a geometry are really conditions about residues of small rank. This motivates the following definition of a diagram.

Let Δ be a finite set. Assume that, for any distinct $i, j \in \Delta$, a class $\mathcal{G}_{ij}$ of geometries of rank 2 is given, whose two types of varieties are called "points" and "blocks". Suppose that the geometries in $\mathcal{G}_{ji}$ are the duals of those in $\mathcal{G}_{ij}$. The set Δ equipped with these collections of geometries is called a *diagram*. It is represented pictorially by taking a "node" for each element of Δ, with an "edge" between each pair of nodes, the edge from i to j being adorned or labelled with some symbol for the class $\mathcal{G}_{ij}$. We will see examples later.

A geometry G belongs to the diagram $(\Delta, (\mathcal{G}_{ij} : i, j \in \Delta))$ if Δ is the set of types of G and, for all distinct $i, j \in \Delta$, and all residues G_F in G with rank 2 and type $\{i, j\}$, G_F is isomorphic to a member of $\mathcal{G}_{ij}$ (where we take points and blocks in G_F to be varieties of types i and j respectively).

In order to illustrate this idea, we need to define some classes of rank 2 geometries to use in diagrams. Some of these we have met already; but the most important is the most trivial: A *digon* is a rank 2 geometry (having at least two points and at least two blocks) in which any point and block are incident; in other words, a complete bipartite graph containing a cycle. By abuse of notation, the "labelled edge" used to represent digons is the absence of an edge! This is done in part because most of the rank 2 residues of our geometries will be digons, and this convention leads to uncluttered pictorial representations of diagrams.

A *partial linear space* is a rank 2 geometry in which two points lie on at most one line (and dually, two lines meet in at most one point). It is represented by an edge with the label Π, thus:



We already met the concepts *linear space* and *generalised projective plane*: they are partial linear spaces in which the first, resp. both, occurrences of "at most" are replaced by "exactly". They are represented by edges with label L and without any label, respectively. (Conveniently, the labels for the self-dual concepts of "partial linear space" and "generalised projective plane" coincide with their mirror-images, while the label for "linear space" does not.) Note that a projective plane is a thick generalised projective plane. Another specialisation of linear spaces, a "circle" or "complete graph", has all lines of cardinality 2; it is denoted by an edge with label c.

Now we can give an example:

Proposition 5.2 *A projective geometry of dimension n has the diagram*

Proof Transversality and residual connectivity are straightforward to check. We verify the rank 2 residues. Take the types to be the dimensions $0,1,\ldots,n-1$, and let F be a flag of cotype $\{i,j\}$, where $i<j$.

Case 1: $j=i+1$. Then F has the form

$$U_0 < U_1 < \ldots < U_{i-1} < U_{i+2} < \ldots < U_{n-1}.$$

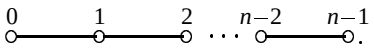
Its residue consists of all subspaces of dimension i or $i+1$ between U_{i-1} and U_{i+2}; this is clearly the projective plane based on the rank 3 vector space U_{i+2}/U_{i-1}.

Case 2: $j>i+1$. Now the flag F looks like

$$U_0 < \ldots < U_{i-1} < U_{i+1} < \ldots < U_{j-1} < U_{j+1} < \ldots < U_{n-1}.$$

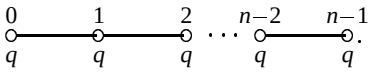
Its residue consists of all subspaces lying either between U_{i-1} and U_{i+1}, or between U_{j-1} and U_{j+1}. Any subspace X of the first type is incident with any subspace Y of the second, since $X < U_{i+1} \leq U_{j-1} < Y$. So the residue is a digon. ∎

In diagrams, it is convenient to label the nodes with the corresponding elements of Δ. For example, in the case of a projective geometry of dimension n, we take the labels to be the dimensions of varieties represented by the nodes, thus:



I will use the convention that labels are placed above the nodes where possible. This reserves the space below the nodes for another use, as follows.

A transversal geometry is said to have *orders*, or *parameters*, if there are numbers s_i (for $i \in \Delta$) with the property that any flag of cotype i is contained in exactly $s_i + 1$ maximal flags. If so, these numbers s_i are the orders (or parameters). Now, if G is a geometry with orders, then G is thick/firm/thin respectively if and only if all orders are $> 1/\geq 1/= 1$ respectively. We will write the orders beneath the nodes, where appropriate. Note that a projective plane of order n (as defined earlier) has orders n, n (in the present terminology). Thus, the geometry $\mathrm{PG}(n,q)$ has diagram



We conclude this section with some general results about Buekenhout geometries. These results depend on our convention that a non-edge symbolises a digon.

Proposition 5.3 *Let the diagram Δ be the disjoint union of Δ_1 and Δ_2, with no edges between these sets. Then a variety with type in Δ_1 and one with type in Δ_2 are incident.*

Proof We use induction on the rank. For rank 2, Δ is the diagram of a digon, and the result is true by definition. So assume that $|\Delta| > 2$, and (without loss of generality) that $|\Delta_1| > 1$.

Let X_i be the set of varieties with type in Δ_i, for $i = 1, 2$. By the inductive hypothesis, if $x, y \in X_1$ with xIy, then $R(x) \cap X_2 = R(y) \cap X_2$. (Considering $R(x)$, we see that every variety in $R(x) \cap X_2$ is incident with y, so the left-hand set is contained in the right-hand set. Reversing the rôles of x and y establishes the result.) Now by connectedness, $R(x) \cap X_2$ is independent of $x \in X_1$. (Note that Proposition 5.1 is being used here.) But this set must be X_2, since every variety in X_2 is incident with some variety in X_1. ■

A diagram is *linear* if the "non-digon" edges form a simple path, as in the diagram for projective spaces in Proposition 5.3 above.

Suppose that one particular type in a geometry is selected, and varieties of that type are called points. Then the *shadow*, or *point-shadow*, of a variety x is the set $\mathrm{Sh}(x)$ of varieties incident with x. Sometimes we write $\mathrm{Sh}_0(x)$, where 0 is the type of a point. In a geometry with a linear diagram, the convention is that points are varieties of the left-most type.

Corollary 5.4 *In a linear diagram, if xIy, and the type of y is further to the right than that of x, then* $\mathrm{Sh}(x) \subseteq \mathrm{Sh}(y)$.

Proof $R(x)$ has disconnected diagram, with points and the type of y in different components; so, by Proposition 5.2, every point in $R(x)$ is incident with y. ■

Exercises

1. (a) Construct a geometry which is connected but not residually connected.
(b) Show that, if G has any of the following properties, then so does any residue of G of rank at least 2: residually connected, transversal, thick, firm, thin.
2. Show that any generalised projective geometry belongs to the diagram

3. (a) A *chamber* of a transversal geometry G is a maximal flag. Let $\mathcal{F}$ be the set of chambers of the geometry G. Form a graph with vertex set $\mathcal{F}$ by joining two chambers which coincide in all but one variety. G is said to be *chamber-connected* if this graph is connected. Prove that a residually connected geometry is chamber-connected, and a chamber-connected geometry is connected.

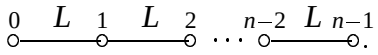
(b) Consider the 3-dimensional affine space $\mathrm{AG}(3,F)$ over the field F. Take three types of varieties: points (type 0), lines (type 1), and parallel classes of planes (type 2). Incidence between points and lines is as usual; a line L and a parallel class C of planes are incident if L lies in some plane of C; and any variety of type 0 is incident with any variety of type 2. Show that this geometry is chamber-connected but not residually connected.

(c) Let V be a six-dimensional vector space over a field F, with a basis $\{e_1, e_2, e_3, f_1, f_2, f_3\}$. Let G be the additive group of V, and let H_1, H_2, H_3 be the additive groups of the three subspaces $\langle e_2, e_3, f_1\rangle$, $\langle e_3, e_1, f_2\rangle$, and $\langle e_1, e_2, f_3\rangle$. Form the *coset geometry* $\mathcal{G}(G,(H_1,H_2,H_3))$: its vaarieties of type i are the cosets of H_i in G, and two varieties are incident if and only if the corresponding cosets have non-empty intersection. Show that this geometry is connected but not chamber-connected.

5.2 Some special diagrams

In this section, we first consider geometries with linear diagram in which all strokes are linear spaces; then we specialise some or all of these linear spaces to projective or affine planes. We will see that the axiomatisations of projective and affine spaces can be expressed very simply in this formalism.

Theorem 5.5 *Let G be a geometry with diagram*



Let varieties of type 0 and 1 be points and lines.

(a) The points and shadows of lines form a linear space $\mathcal{L}$.

(b) The shadow of any variety is a subspace of $\mathcal{L}$.

(c) $\mathrm{Sh}_0(x) \subseteq \mathrm{Sh}_0(y)$ if and only if x is incident with y.

(d) If x is a variety and p a point not incident with x, then there is a unique variety y incident with x and p such that $\tau(y) = \tau(x) + 1$.

Proof (a) We show that two points lie on at least one line by induction on the rank. There is a path between any two points using only points and lines, by Proposition 5.2; so it suffices to show that any such path of length greater than 2 can be shortened. So assume $pILIqIMIr$, where p,q,r are points and L,M lines. By the induction hypothesis, the POINTs L and M of $R(q)$ lie in a LINE Π, a plane of G incident with L and M. By Corollary 5.4, p and q are incident with Π. Since Π is a linear space, there is a line through p and q. (The convention of using capitals for varieties in $R(q)$ is used here.)

Now suppose that two lines L and M contain the two points p and q. Considering $R(p)$, we find a plane Π incident with L and M and hence with p and q. But Π is a linear space, so $L = M$.

(b) Let y be any variety, and $p,q \in \mathrm{Sh}_0(y)$. Since points and lines incident with y form a linear space by (a), there is a line incident with p,q and y. This must be the unique line incident with p and q; and, by Corollary 5.4, all its points are incident with y and so are in $\mathrm{Sh}_0(y)$.

(c) The reverse implication is Corollary 5.4. So suppose that $\mathrm{Sh}_0(x) \subseteq \mathrm{Sh}_0(y)$. Take $p \in \mathrm{Sh}_0(x)$. Then, in $R(p)$, we have $\mathrm{Sh}_1(x) \subseteq \mathrm{Sh}_1(y)$ (since these shadows are linear subspaces) , and so xIy by induction. (The base case of the induction, where x is a line, is covered by (b).)

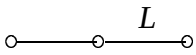
(d) This is clear if x is a point. Otherwise, choose $q \in \mathrm{Sh}_0(x)$, and apply induction in $R(q)$ (replacing p by the line pq). ■

Theorem 5.6 *A geometry with diagram*

is a generalised projective space (of finite dimension).

Proof By Theorem 5.5(d), a potential Veblen configuration lies in a plane; since planes are projective, Veblen's axiom holds. It remains to show that every linear subspace is the shadow of some variety; this follows easily by induction. ■

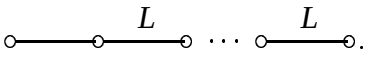
Theorem 5.7 *A geometry with diagram*



consists of the points, lines and planes of a (possibly infinite-dimensional) generalised projective space.

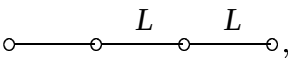
Proof Veblen's axiom is verified as in Theorem 5.6. It is clear that every point, line or plane corresponds to a variety. ■

Remark. Consider geometries with the diagram



By the argument for Theorem 5.7, we have all the points, lines and planes, and some higher-dimensional varieties, of a generalised projective space. Examples arise by taking all the flats of dimension at most $r-1$, where r is the rank. However, there are other examples. A simple case, with $r = 4$, can be constructed as follows.

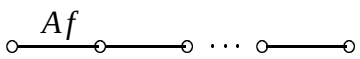
Let $\mathcal{P}$ be a projective space of countable dimension over a finite field F. Enumerate the 3-dimensional and 4-dimensional subspaces in lists $T_0, T_1, \ldots$ and $F_0, F_1, \ldots$. Now construct a set $\mathcal{F}$ of 4-dimensional subspaces in stages as follows. At the n^{th} stage, if T_n is already contained in a member of $\mathcal{F}$, do nothing. Otherwise, of the infinitely many subspaces F_j which contain T_n, only finitely many are excluded because they contain any T_m with $m < n$; let F_i be the one with smallest index which is not excluded, and adjoin it to $\mathcal{F}$. At the conclusion, any 3-dimensional subspace is contained in a unique member of $\mathcal{F}$. Then the points, lines, planes, and subspaces in $\mathcal{F}$ form a geometry with the diagram



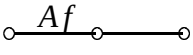
where the first L denotes the points and lines in 3-dimensional projective space over F.

Now we turn to affine spaces, where similar results hold. The label Af on a stroke will denote the class of affine planes.

Theorem 5.8 *A geometry with diagram*



is an affine space of finite dimension.

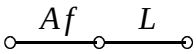
Proof It is a linear space whose planes are affine (that is, satisfying condition (AS1) of Section 11). We must show that parallelism is transitive. So suppose that $L_1 \| L_2 \| L_3$, but $L_1 \not\| L_3$. Then all three lines lie in a subspace of dimension 3; so it is enough to deduce a contradiction in the case of geometries of rank 3. Note that, for a geometry with diagram 

, two planes which have a common point must meet in a line.

Let Π_1 be the plane through L_1 and L_2, and Π_2 the plane through p and L_3, where p is a point of L_1. Then Π_1 and Π_2 both contain p, so they meet in a line $M \neq L_1$. Then M is not parallel to L_2, so meets it in a point q, But then Π_2 contains L_3 and q, hence L_2, and so is equal to Π_1, a contradiction.

The fact that all linear subspaces are shadows of varieties is proved as in Theorem 5.6. ■

Theorem 5.9 *A geometry with diagram*

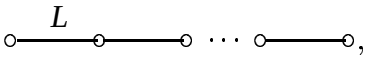


in which some line has more than three points, consists of the points, lines and planes of a (possibly infinite-dimensional) affine space.

The proof is as for Theorem 5.7, using Buekenhout's Theorem 3.10. ■

Exercises

1. Consider a geometry of rank n with diagram



,

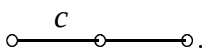
in which all lines have the same finite cardinality k, and all the projective planes have the same finite order q.

(a) If $n \geq 4$, prove that the geometry is either projective ($q = k-1$) or affine ($q = k$).

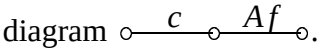
(b) If $n = 3$, prove that $q = k-1, k, k^2$ or $k(k^2+1)$.

(This result is due to Doyen and Hubaut [16]).

2. Construct an infinite "free-like" geometry with diagram



(Ensure that three points lie in a unique plane, while two planes meet in two points.)

3. (a) Show that an inversive plane belongs to the diagram 

. What are the varieties?

(b) Show how to construct a geometry with diagram

(n nodes) from an ovoid in $\mathrm{PG}(n, F)$ (see Section 4.4).

6

Polar spaces

Now we begin on our second major theme, polar spaces. This chapter corresponds to the first half of Chapter 1, and gives the algebraic description of polar spaces. The algebraic background required is more elaborate (vector spaces with forms, rather than just vector spaces), accounting for the increased length. The first section, on polarities of projective spaces, provides motivation for the introduction of the (Hermitian and quadratic) forms.

6.1 Dualities and polarities

Recall that the dual V^* of a finite-dimensional (left) vector space V over a skew field F can be regarded as a left vector space of the same dimension over the opposite field F°, and there is thus an inclusion-reversing bijection between the projective spaces $\mathrm{PG}(n,F)$ and $\mathrm{PG}(n,F^\circ)$. If it happens that F and F° are isomorphic, then there exists a *duality* of $\mathrm{PG}(n,F)$, an inclusion-reversing bijection of $\mathrm{PG}(n,F)$.

Conversely, if $\mathrm{PG}(n,F)$ admits a duality (for $n>1$), then F is isomorphic to F°, as follows from the FTPG (see Section 1.3). We will examine this conclusion and make it more detailed.

So let π be a duality of $\mathrm{PG}(n,F)$, $n>1$. Composing π with the natural isomorphism from $\mathrm{PG}(n,F)$ to $\mathrm{PG}(n,F^\circ)$, we obtain an inclusion-preserving bijection θ from $\mathrm{PG}(n,F)$ to $\mathrm{PG}(n,F^\circ)$. According to the FTPG, θ is induced by a semilinear transformation T from $V=F^{n+1}$ to its dual space V^*, associated with an isomorphism $\sigma : F \to F^\circ$, which can be regarded as being an anti-automorphism of F:

that is,

$$\begin{aligned}(\mathbf{v}_1+\mathbf{v}_2)T &= \mathbf{v}_1T+\mathbf{v}_2T,\\ (\alpha\mathbf{v})T &= \alpha^\sigma\mathbf{v}T.\end{aligned}$$

Define a function $b : V \times V \to F$ by the rule

$$b(\mathbf{v},\mathbf{w}) = (\mathbf{v})(\mathbf{w}T),$$

that is, the result of applying the element $\mathbf{w}T$ of V^* to $\mathbf{v}$. Then b is a *sesquilinear form*: it is linear as a function of the first argument, and semilinear as a function of the second — this means that

$$b(\mathbf{v},\mathbf{w}_1+\mathbf{w}_2) = b(\mathbf{v},\mathbf{w}_1)+b(\mathbf{v},\mathbf{w}_2)$$

and

$$b(\mathbf{v},\alpha\mathbf{w}) = \alpha^\sigma b(\mathbf{v},\mathbf{w}).$$

(The prefix "sesqui-" means "one-and-a-half".) If we need to emphasise the anti-automorphism σ, we say that b is σ-sesquilinear. If σ is the identity, then the form is *bilinear*.

The form b is also *non-degenerate*, in the sense that

$$(\forall \mathbf{w}\in V)(b(\mathbf{v},\mathbf{w})=0 \quad\Rightarrow\quad \mathbf{v}=0$$

and

$$(\forall \mathbf{v}\in V)(b(\mathbf{v},\mathbf{w})=0 \quad\Rightarrow\quad \mathbf{w}=0.$$

(The second condition asserts that T is one-to-one, so that if $\mathbf{w}\neq 0$ then $\mathbf{w}T$ is a non-zero functional. The first asserts that T is onto: only the zero vector is annihilated by every functional in the dual space.)

So, we have:

Theorem 6.1 *Any duality of* $\mathrm{PG}(n,F)$*, for* $n>1$*, is induced by a non-degenerate* σ*-sesquilinear form on the underlying vector space, where* σ *is an anti-automorphism of* F*.* ■

Conversely, any non-degenerate sesquilinear form on V induces a duality. We can short-circuit the passage to the dual space, and write the duality as

$$U \mapsto U^\perp = \{\mathbf{v}\in V : b(\mathbf{v},\mathbf{w})=0 \text{ for all } \mathbf{w}\in U\}.$$

Obviously, a duality applied twice is a collineation. The most important types of dualities are those whose square is the identity. A *polarity* of $\mathrm{PG}(n,F)$is a duality $\perp$ which satisfies $U^{\perp\perp} = U$ for all flats U of $\mathrm{PG}(n,F)$.

It is a bit difficult to motivate the detailed study of polarities at this stage; but it will turn out that they give rise to a class of geometries (the polar spaces) with properties similar to those of projective spaces. To put it somewhat vaguely, we are trying to add some extra structure to a projective space; if a duality is not a polarity, then its square is a non-identity collineation, and some of the extra structure arises from this collineation. Only in the case of a polarity is the extra structure "primitive".

A sesquilinear form b is *reflexive* if $b(\mathbf{v},\mathbf{w}) = 0$ implies $b(\mathbf{w},\mathbf{v}) = 0$.

Proposition 6.2 *A duality is a polarity if and only if the sesquilinear form defining it is reflexive.*

Proof b is reflexive if and only if

$$\mathbf{v} \in \langle \mathbf{w} \rangle^{\perp} \Rightarrow \mathbf{w} \in \langle \mathbf{v} \rangle^{\perp}.$$

Hence, if b is reflexive, then $U \subseteq U^{\perp\perp}$ for all subspaces U. But by non-degeneracy, $\dim U^{\perp\perp} = \dim V - \dim U^{\perp} = \dim U$; and so $U = U^{\perp\perp}$ for all U. Conversely, given a polarity $\perp$, if $\mathbf{w} \in \langle \mathbf{v} \rangle^{\perp}$, then $\mathbf{v} \in \langle \mathbf{v} \rangle^{\perp\perp} \subseteq \langle \mathbf{w} \rangle^{\perp}$ (since inclusions are reversed). ■

We now turn to the classification of reflexive forms. For convenience, from now on F will always be assumed to be commutative. (Note that, if the anti-automorphism σ is an automorphism, and in particular if σ is the identity, then F is automatically commutative.)

The form b is said to be σ-*Hermitian* if $b(\mathbf{w},\mathbf{v}) = b(\mathbf{v},\mathbf{w})^{\sigma}$ for all $\mathbf{v},\mathbf{w} \in V$. This implies that, for any $\mathbf{v}$, $b(\mathbf{v},\mathbf{v})$ lies in the fixed field of σ. If σ is the identity, such a form (which is bilinear) is called *symmetric*.

A bilinear form b is called *alternating* if $b(\mathbf{v},\mathbf{v}) = 0$ for all $\mathbf{v} \in V$. This implies that $b(\mathbf{w},\mathbf{v}) = -b(\mathbf{v},\mathbf{w})$ for all $\mathbf{v},\mathbf{w} \in V$. (Expand $b(\mathbf{v}+\mathbf{w},\mathbf{v}+\mathbf{w}) = 0$, and note that two of the four terms are zero.) Hence, if the characteristic is 2, then any alternating form is symmetric (but not conversely); but, in characteristic different from 2, only the zero form is both symmetric and alternating.

Clearly, an alternating or Hermitian form is reflexive. Conversely, we have the following:

Theorem 6.3 *A non-degenerate reflexive σ-sesquilinear form is either alternating, or a scalar multiple of a σ-Hermitian form. In the latter case, if σ is the identity, then the scalar can be taken to be 1.*

I will not give the complete proof of this theorem. The next result shows that $\sigma^2 = 1$, and then the proof of the theorem is given in the case of a bilinear form (that is, when $\sigma = 1$).

Proposition 6.4 *If b is a non-zero reflexive σ-sesquilinear form, then σ^2 is the identity.*

Proof Note first that a form is σ-sesquilinear if and only if it is additive in each variable and satisfies

$$b(\alpha\mathbf{v},\mathbf{w}) = \alpha b(\mathbf{v},\mathbf{w}), \qquad b(\mathbf{v},\beta\mathbf{w}) = b(\mathbf{v},\mathbf{w})\beta^\sigma.$$

Step 1 If b is alternating, then $\sigma = 1$. For we can choose $\mathbf{v}$ and $\mathbf{w}$ with $b(\mathbf{v},\mathbf{w}) = -b(\mathbf{w},\mathbf{v}) = 1$. Then for any $\alpha \in F$, we have

$$\begin{aligned} \alpha &= \alpha b(\mathbf{v},\mathbf{w}) \\ &= b(\alpha\mathbf{v},\mathbf{w}) \\ &= -b(\mathbf{w},\alpha\mathbf{v}) \\ &= -b(\mathbf{w},\mathbf{v})\alpha^\sigma \\ &= \alpha^\sigma. \end{aligned}$$

(Note that this step does not require non-degeneracy, merely that b is not identically zero.)

So we can assume that there exists $\mathbf{v}$ with $b(\mathbf{v},\mathbf{v}) \neq 0$. Multiplying b by a non-zero scalar (this does not affect the hypotheses), we may assume that $b(\mathbf{v},\mathbf{v}) = 1$.

Step 2 Assume for a contradiction that $\sigma^2 \neq 1$. For any vector $\mathbf{w}$, if $b(\mathbf{w},\mathbf{v}) \neq 0$, then we can replace $\mathbf{w}$ by its product with a non-zero scalar to assume $b(\mathbf{w},\mathbf{v}) = 1$. Then $b(\mathbf{w}-\mathbf{v},\mathbf{v}) = 0$, and so $b(\mathbf{v},\mathbf{w}-\mathbf{v}) = 0$, whence $b(\mathbf{v},\mathbf{w}) = 1$. We *claim* that $b(\mathbf{w},\mathbf{w}) = 1$.

Proof Suppose that $\alpha = b(\mathbf{w},\mathbf{w}) \neq 1$. Note first that $b(\mathbf{w}-\alpha\mathbf{v},\mathbf{v}) = 0$, and so $b(\mathbf{w},\mathbf{w}-\alpha\mathbf{v}) = 0$, whence $\alpha = \alpha^\sigma$. Take any element $\lambda \in F$ with $\lambda \neq 1$, and choose $\mu \in F$ such that $\mu^\sigma = (1-\lambda)^{-1}(\alpha-\lambda)$. Since $\alpha \neq 1$, we have $\mu \neq 1$; and

$$\mu^\sigma - \lambda\mu^\sigma = \alpha - \lambda.$$

This implies, first, that $\lambda = (\alpha-\mu^\sigma)(1-\mu^\sigma)^{-1}$, and second that

$$b(\mathbf{w}-\lambda\mathbf{v},\mathbf{w}-\mu\mathbf{v}) = \alpha - \lambda - \mu^\sigma + \lambda\mu^\sigma = 0.$$

Hence $b(\mathbf{w}-\mu\mathbf{v},\mathbf{w}-\lambda\mathbf{v}) = 0$, and we obtain

$$\alpha - \mu - \lambda^\sigma + \mu\lambda^\sigma = 0.$$

Applying σ to this equation and using the fact that $\alpha^\sigma = \alpha$, we obtain

$$\alpha - \mu^\sigma - \lambda^{\sigma^2} + \lambda^{\sigma^2}\mu^\sigma = 0,$$

whence

$$\lambda^{\sigma^2} = (\alpha-\mu^\sigma)(1-\mu^\sigma)^{-1} = \lambda.$$

But λ was an arbitrary element different from 1. Since clearly $1^\sigma = 1$, we have $\sigma^2 = 1$, contrary to assumption.

Step 3 Let $W = \mathbf{v}^\perp$. Then $V = \langle\mathbf{v}\rangle \oplus W$, and $\mathrm{rk}(W) \geq 1$. For any $\mathbf{x} \in W$, we have $b(\mathbf{v},\mathbf{v}) = b(\mathbf{v}+\mathbf{x},\mathbf{v}) = 1$, and so by Step 2, we have $b(\mathbf{v}+\mathbf{x},\mathbf{v}+\mathbf{x}) = 1$. Thus $b(\mathbf{x},\mathbf{x}) = -2$. Putting $x = 0$, we see that F must have characteristic 2, and that $b|W$ is alternating. But then Step 1 shows that $b|W$ is identically zero, whence W is contained in the radical of b, contrary to the assumed non-degeneracy.

Proof of Theorem 6.3 We have

$$b(\mathbf{u},\mathbf{v})b(\mathbf{u},\mathbf{w}) - b(\mathbf{u},\mathbf{w})b(\mathbf{u},\mathbf{v}) = 0$$

by commutativity; that is, using bilinearity,

$$b(\mathbf{u}, b(\mathbf{u},\mathbf{v})\mathbf{w} - b(\mathbf{u},\mathbf{w})\mathbf{v}) = 0.$$

By reflexivity,

$$b(b(\mathbf{u},\mathbf{v})\mathbf{w} - b(\mathbf{u},\mathbf{w})\mathbf{v}, \mathbf{u}) = 0,$$

whence bilinearity again gives

$$b(\mathbf{u},\mathbf{v})b(\mathbf{w},\mathbf{u}) = b(\mathbf{u},\mathbf{w})b(\mathbf{v},\mathbf{u}). \tag{6.1}$$

Call a vector $\mathbf{u}$ *good* if $b(\mathbf{u},\mathbf{v}) = b(\mathbf{v},\mathbf{u}) \neq 0$ for some $\mathbf{v}$. By (6.1), if $\mathbf{u}$ is good, then $b(\mathbf{u},\mathbf{w}) = b(\mathbf{w},\mathbf{u})$ for all $\mathbf{w}$. Also, if $\mathbf{u}$ is good and $b(\mathbf{u},\mathbf{v}) \neq 0$, then $\mathbf{v}$ is good. But, given any two non-zero vectors $\mathbf{u}_1,\mathbf{u}_2$, there exists $\mathbf{v}$ with $b(\mathbf{u}_i,\mathbf{v}) \neq 0$ for $i = 1,2$. (For there exist $\mathbf{v}_1,\mathbf{v}_2$ with $b(\mathbf{u}_i,\mathbf{v}_i) \neq 0$ for $i = 1,2$, by non-degeneracy; and at least one of $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_1+\mathbf{v}_2$ has the required property.) So, if some vector is good, then every non-zero vector is good, and b is symmetric.

But, putting $\mathbf{u} = \mathbf{w}$ in (6.1) gives

$$b(\mathbf{u},\mathbf{u})(b(\mathbf{u},\mathbf{v}) - b(\mathbf{v},\mathbf{u})) = 0$$

for all $\mathbf{u},\mathbf{v}$. So, if $\mathbf{u}$ is not good, then $b(\mathbf{u},\mathbf{u}) = 0$; and, if no vector is good, then b is alternating. ■

In the next few sections, we develop this theme further.

Exercises

1. Let b be a sesquilinear form on V. Define the *left* and *right radicals* of b to be the subsets

$$\{\mathbf{v} \in V : (\forall \mathbf{w} \in V) b(\mathbf{v},\mathbf{w}) = 0\}$$

and

$$\{\mathbf{v} \in V : (\forall \mathbf{w} \in V) b(\mathbf{w},\mathbf{v}) = 0\}$$

respectively. Prove that the left and right radicals are subspaces of the same rank (if V has finite rank).

(Note: If the left and right radicals are equal, this subspace is called the *radical* of b. This holds if b is reflexive.)

2. Give an example of a bilinear form on an infinite-rank vector space whose left radical is zero and whose right radical is non-zero.

3. Let σ be a (non-identity) automorphism of F of order 2. Let E be the subfield $\mathrm{Fix}(\sigma)$.

(a) Prove that F is of degree 2 over E, i.e., a rank 2 E-vector space.

[See any textbook on Galois theory. Alternately, argue as follows: Take $\lambda \in F \setminus E$. Then λ is quadratic over E, so $E(\lambda)$ has degree 2 over E. Now $E(\lambda)$ contains an element ω such that $\omega^\sigma = -\omega$ (if the characteristic is not 2) or $\omega\sigma =$

$\omega+1$ (if the characteristic is 2). Now, given two such elements, their quotient or difference respectively is fixed by σ, so lies in E.]

(b) Prove that

$$\{\lambda \in F : \lambda\lambda^\sigma = 1\} = \{\varepsilon/\varepsilon^\sigma : \varepsilon \in F\}.$$

[The left-hand set clearly contains the right. For the reverse inclusion, separate into cases according as the characteristic is 2 or not.

If the characteristic is not 2, then we can take $F = E(\omega)$, where $\omega^2 = \alpha \in E$ and $\omega^\sigma = -\omega$. If $\lambda = 1$, then take $\varepsilon = 1$; otherwise, if $\lambda = a + b\omega$, take $\varepsilon = b\alpha + (a-1)\omega$.

If the characteristic is 2, show that we can take $F = E(\omega)$, where $\omega^2+\omega+\alpha = 0$, $\alpha \in E$, and $\omega^\sigma = \omega+1$. Again, if $\lambda = 1$, set $\varepsilon = 1$; else, if $\lambda = a + b\omega$, take $\varepsilon = (a+1) + b\omega$.]

4. Use the result of Exercise 3 to complete the proof of Theorem 6.3 in general.

[If $b(\mathbf{u},\mathbf{u}) = 0$ for all $\mathbf{u}$, the form b is alternating and *bilinear*. If not, suppose that $b(\mathbf{u},\mathbf{u}) \neq 0$ and let $b(\mathbf{u},\mathbf{u})^\sigma = \lambda b(\mathbf{u},\mathbf{u})$. Choosing ε as in Exercise 2 and renormalising b, show that we may assume that $\lambda = 1$, and (with this choice) that b is Hermitian.]

6.2 Hermitian and quadratic forms

We now change ground slightly from the last section. On the one hand, we restrict things by excluding some bilinear forms from the discussion; on the other, we introduce quadratic forms. The loss and gain exactly balance if the characteristic is not 2; but, in characteristic 2, we make a net gain.

Let σ be an automorphism of the commutative field F, of order dividing 2. Let $\mathrm{Fix}(\sigma) = \{\lambda \in F : \lambda^\sigma = \lambda\}$ be the *fixed field* of σ, and $\mathrm{Tr}(\sigma) = \{\lambda + \lambda^\sigma : \lambda \in F\}$ the *trace* of σ. Since σ^2 is the identity, it is clear that $\mathrm{Fix}(\sigma) \supseteq \mathrm{Tr}(\sigma)$. Moreover, if σ is the identity, then $\mathrm{Fix}(\sigma) = F$, and

$$\mathrm{Tr}(\sigma) = \begin{cases} 0 & \text{if } F \text{ has characteristic 2,} \\ F & \text{otherwise.} \end{cases}$$

Let b be a σ-Hermitian form. We observed in the last section that $b(\mathbf{v},\mathbf{v}) \in \mathrm{Fix}(\sigma)$ for all $\mathbf{v} \in V$. We call the form b *trace-valued* if $b(\mathbf{v},\mathbf{v}) \in \mathrm{Tr}(\sigma)$ for all $\mathbf{v} \in V$.

Proposition 6.5 *We have* $\mathrm{Tr}(\sigma) = \mathrm{Fix}(\sigma)$ *unless the characteristic of* F *is 2 and* σ *is the identity.*

Proof $E = \mathrm{Fix}(\sigma)$ is a field, and $K = \mathrm{Tr}(\sigma)$ is an E-vector space contained in E (Exercise 1). So, if $K \neq E$, then $K = 0$, and σ is the map $x \mapsto -x$. But, since σ is a field automorphism, this implies that the characteristic is 2 and σ is the identity.

Thus, in characteristic 2, symmetric bilinear forms which are not alternating are not trace-valued; but this is the only obstruction. We introduce quadratic forms to repair this damage. But, of course, quadratic forms can be defined in any characteristic. However, we note at this point that Proposition 6.5 depends in a crucial way on the commutativity of F; this leaves open the possibility of additional types of polar spaces defined by so-called *pseudoquadratic forms*. These will be discussed briefly in Section 7.6.

Let V be a vector space over F. A *quadratic form* on V is a function $f : V \to F$ satisfying

- $f(\lambda \mathbf{v}) = \lambda^2 f(\mathbf{v})$ for all $\lambda \in F$, $\mathbf{v} \in V$;
- $f(\mathbf{v}+\mathbf{w}) = f(\mathbf{v}) + f(\mathbf{w}) + b(\mathbf{v},\mathbf{w})$, where b is bilinear.

Now, if the characteristic of F is not 2, then b is a symmetric bilinear form. Each of f and b determines the other, by

$$b(\mathbf{v},\mathbf{w}) = f(\mathbf{v}+\mathbf{w}) - f(\mathbf{v}) - f(\mathbf{w})$$

and

$$f(\mathbf{v}) = \tfrac{1}{2} b(\mathbf{v},\mathbf{v}),$$

the latter equation coming from the substitution $\mathbf{v} = \mathbf{w}$ in the second defining condition. So nothing new is obtained.

On the other hand, if the characteristic of F is 2, then b is an alternating bilinear form, and f cannot be recovered from b. Indeed, many different quadratic forms correspond to the same bilinear form. (Note that the quadratic form does give extra structure to the vector space; we'll see that this structure is geometrically similar to that provided by an alternating or Hermitian form.)

We say that the bilinear form is obtained by *polarisation* of f.

Now let b be a symmetric bilinear form over a field of characteristic 2, which is not alternating. Set $f(\mathbf{v}) = b(\mathbf{v},\mathbf{v})$. Then we have

$$f(\lambda \mathbf{v}) = \lambda^2 f(\mathbf{v})$$

and

$$f(\mathbf{v}+\mathbf{w}) = f(\mathbf{v}) + f(\mathbf{w}),$$

since $b(\mathbf{v},\mathbf{w})+b(\mathbf{w},\mathbf{v})=0$. Thus f is "almost" a semilinear form; the map $\lambda\mapsto\lambda^2$ is a homomorphism of the field F with kernel 0, but it may fail to be an automorphism. But in any case, the kernel of f is a subspace of V, and the restriction of b to this subspace is an alternating bilinear form. So again, in the spirit of the vague comment motivating the study of polarities in the last section, the structure provided by the form b is not "primitive". For this reason, we do not consider symmetric bilinear forms in characteristic 2 at all. However, as indicated above, we will consider quadratic forms in characteristic 2.

Now, in characteristic different from 2, we can take either quadratic forms or symmetric bilinear forms, since the structural content is the same. For consistency, we will take quadratic forms in this case too. This leaves us with three "types" of forms to study: alternating bilinear forms; σ-Hermitian forms where σ is not the identity; and quadratic forms.

We have to define the analogue of non-degeneracy for quadratic forms. Of course, we could require that the bilinear form obtained by polarisation is non-degenerate; but this is too restrictive. We say that a quadratic form f is *non-singular* if

$$(f(\mathbf{v})=0\ \&\ (\forall\mathbf{w}\in V)b(\mathbf{v},\mathbf{w})=0)\quad\Rightarrow\quad\mathbf{v}=0$$

where b is the associated bilinear form; that is, if the form f is non-zero on every non-zero vector of the radical.

If the characteristic is not 2, then non-singularity is equivalent to non-degeneracy of the bilinear form.

Now suppose that the characteristic is 2, and let W be the radical. Then b is identically zero on W; so the restriction of f to W satisfies

$$\begin{aligned} f(\mathbf{v}+\mathbf{w}) &= f(\mathbf{v})+f(\mathbf{w}),\\ f(\lambda\mathbf{v}) &= \lambda^2 f(\mathbf{v}).\end{aligned}$$

As above, f is very nearly semilinear. The field F is called *perfect* if every element is a square. In this case, f is indeed semilinear, and its kernel is a hyperplane of W. We conclude:

Theorem 6.6 *Let f be a non-singular quadratic form, which polarises to b, over a field F.*

(a) If the characteristic of F is not 2, then b is non-degenerate.

(b) If F is a perfect field of characteristic 2, then the radical of b has rank at most 1.

Exercises

1. Let σ be an automorphism of a commutative field F such that σ^2 is the identity.

(a) Prove that $\mathrm{Fix}(\sigma)$ is a subfield of F.

(b) Prove that $\mathrm{Tr}(\sigma)$ is closed under addition, and under multiplication by elements of $\mathrm{Fix}(\sigma)$.

2. Let b be an alternating bilinear form on a vector space V over a field F of characteristic 2. Let $(\mathbf{v}_i : i \in I)$ be a basis for V, and q any function from I to F. Show that there is a unique quadratic form with the properties that $f(\mathbf{v}_i) = q(i)$ for every $i \in I$, and f polarises to b.

3. (a) Construct an imperfect field of characteristic 2.

(b) Construct a non-singular quadratic form with the property that the radical of the associated bilinear form has rank greater than 1.

4. Show that finite fields of characteristic 2 are perfect. (Hint: the multiplicative group is cyclic of odd order.)

6.3 Classification of forms

As explained in the last section, we now consider a vector space V of finite rank equipped with a form of one of the following types: a non-degenerate alternating bilinear form b; a non-degenerate σ-Hermitian form b, where σ is not the identity; or a non-singular quadratic form f. In the third case, we let b be the bilinear form obtained by polarising f; then b is alternating or symmetric according as the characteristic is or is not 2, but b may be degenerate. In the other two cases, we define a function $f : V \to F$ defined by $f(\mathbf{v}) = b(\mathbf{v}, \mathbf{v})$ — this is identically zero if b is alternating. See Exercise 1 for the Hermitian case.

We say that V is *anisotropic* if $f(\mathbf{v}) \neq 0$ for all $\mathbf{v} \neq 0$. Also, V is a *hyperbolic line* if it is spanned by vectors $\mathbf{v}$ and $\mathbf{w}$ with $f(\mathbf{v}) = f(\mathbf{w}) = 0$ and $b(\mathbf{v}, \mathbf{w}) = 1$. (The vectors $\mathbf{v}$ and $\mathbf{w}$ are linearly independent, so V has rank 2; so, projectively, it is a "line".)

Theorem 6.7 *A space carrying a form of one of the above types is the direct sum of a number r of hyperbolic lines and an anisotropic space U. The number r and the isomorphism type of U are invariants of V.*

Proof If V is anisotropic, then there is nothing to prove. (V cannot contain a hyperbolic line.) So suppose that V contains a vector $\mathbf{v} \neq 0$ with $f(\mathbf{v}) = 0$.

We claim that there is a vector $\mathbf{w}$ with $b(\mathbf{v},\mathbf{w}) \neq 0$. In the alternating and Hermitian cases, this follows immediately from the non-degeneracy of the form. In the quadratic case, if no such vector exists, then $\mathbf{v}$ is in the radical of b; but $\mathbf{v}$ is a singular vector, contradicting the non-singularity of f.

Multiplying $\mathbf{w}$ by a non-zero constant, we may assume that $b(\mathbf{v},\mathbf{w}) = 1$.

Now, for any value of λ, we have $b(\mathbf{v},\mathbf{w}-\lambda\mathbf{v}) = 1$. We wish to choose λ so that $f(\mathbf{w}-\lambda\mathbf{v}) = 0$; then $\mathbf{v}$ and $\mathbf{w}$ will span a hyperbolic line. Now we distinguish cases. If b is alternating, then any value of λ works. If b is Hermitian, we have

$$\begin{aligned} f(\mathbf{w}-\lambda\mathbf{v}) &= f(\mathbf{w}) - \lambda b(\mathbf{v},\mathbf{w}) - \lambda^\sigma b(\mathbf{w},\mathbf{v}) + \lambda\lambda^\sigma f(\mathbf{v}) \\ &= f(\mathbf{w}) - (\lambda + \lambda^\sigma); \end{aligned}$$

and, since b is trace-valued, there exists λ with $\mathrm{Tr}(\lambda) = f(\mathbf{w})$. Finally, if f is quadratic, we have

$$\begin{aligned} f(\mathbf{w}-\lambda\mathbf{v}) &= f(\mathbf{w}) - \lambda b(\mathbf{w},\mathbf{v}) + \lambda^2 f(\mathbf{v}) \\ &= f(\mathbf{w}) - \lambda, \end{aligned}$$

so we choose $\lambda = f(\mathbf{w})$.

Now let W_1 be the hyperbolic line $\langle \mathbf{v}, \mathbf{w}-\lambda\mathbf{v}\rangle$, and let $V_1 = W_1^\perp$, where orthogonality is defined with respect to the form b. It is easily checked that $V = V_1 \oplus W_1$, and the restriction of the form to V_1 is still non-degenerate or non-singular, as appropriate. Now the existence of the decomposition follows by induction.

I will omit the proof of uniqueness. ■

The number r of hyperbolic lines is called the *polar rank* or *Witt index* of V. I do not know of a commonly accepted term for U; I will call it the *germ* of V, for reasons which will become clear shortly.

To complete the classification of forms over a given field, it is necessary to determine all the anisotropic spaces. In general, this is not possible; for example, the study of positive definite quadratic forms over the rational numbers leads quickly into deep number-theoretic waters. I will consider the cases of the real and complex numbers and finite fields.

First, though, the alternating case is trivial:

Proposition 6.8 *The only anisotropic space carrying an alternating bilinear form is the zero space.* ■

In combination with Theorem 6.7, this shows that a space carrying a non-degenerate alternating bilinear form is a direct sum of hyperbolic lines.

Over the real numbers, Sylvester's theorem asserts that any quadratic form in n variables is equivalent to the form

$$x_1^2 + \ldots + x_r^2 - x_{r+1}^2 - \ldots - x_{r+s}^2,$$

for some r, s with $r + s \leq n$. If the form is non-singular, then $r + s = n$. If both r and s are non-zero, there is a non-zero singular vector (with 1 in positions 1 and $r+1$, 0 elsewhere). So we have:

Proposition 6.9 *If V is a real vector space of rank n, then an anisotropic form on V is either positive definite or negative definite; there is a unique form of each type up to invertible linear transformation, one the negative of the other.* ■

The reals have no non-identity automorphisms, so Hermitian forms do not arise.

Over the complex numbers, the following facts are easily shown:

(a) There is a unique non-singular quadratic form (up to equivalence) in n variables for any n. A space carrying such a form is anisotropic if and only if $n \leq 1$.

(b) If σ denotes complex conjugation, the situation for σ-Hermitian forms is the same as for quadratic forms over the reals: anisotropic forms are positive or negative definite, and there is a unique form of each type, one the negative of the other.

For finite fields, the position is as follows.

Theorem 6.10 *(a) An anisotropic quadratic form in n variables over $\mathrm{GF}(q)$ exists if and only if $n \leq 2$. There is a unique form for each n except when $n = 1$ and q is odd, in which case there are two forms, one a non-square multiple of the other.*

(b) Let $q = r^2$ and let σ be the field automorphism $\alpha \mapsto \alpha^r$. Then there is an anisotropic σ-Hermitian form in n variables if and only if $n \leq 1$. The form is unique in each case.

Proof (a) Consider first the case where the characteristic is not 2. The multiplicative group of $\mathrm{GF}(q)$ is cyclic of even order $q-1$; so the squares form a subgroup of index 2, and if η is a fixed non-square, then every non-square has the form $\eta\alpha^2$ for some α. It follows easily that any quadratic form in one variable is equivalent to either x^2 or ηx^2.

Next, consider non-singular forms in two variables. By completing the square, such a form is equivalent to one of x^2+y^2, $x^2+\eta y^2$, $\eta x^2+\eta y^2$.

Suppose first that $q \equiv 1 \pmod 4$. Then -1 is a square, say $-1=\beta^2$. (In the multiplicative group, -1 has order 2, so lies in the subgroup of even order $\frac{1}{2}(q-1)$ consisting of squares.) Thus $x^2+y^2=(x+\beta y)(x-\beta y)$, and the first and third forms are not anisotropic. Moreover, any form in 3 or more variables, when converted to diagonal form, contains one of these two, and so is not anisotropic either.

Now consider the other case, $q \equiv -1 \pmod 4$. Then -1 is a non-square (since the group of squares has odd order), so the second form is $(x+y)(x-y)$, and is not anisotropic. Moreover, the set of squares is not closed under addition (else it would be a subgroup of the additive group, but $\frac{1}{2}(q+1)$ doesn't divide q); so there exist two squares whose sum is a non-square. Multiplying by a suitable square, there exist β,γ with $\beta^2+\gamma^2=-1$. Then

$$-(x^2+y^2)=(\beta x+\gamma y)^2+(\gamma x-\beta y)^2,$$

and the first and third forms are equivalent. Moreover, a form in three variables is certainly not anisotropic unless it is equivalent to $x^2+y^2+z^2$, and this form vanishes at the vector $(\beta,\gamma,1)$; hence there is no anisotropic form in three or more variables.

The characteristic 2 case is an exercise (see Exercise 3).

(b) Now consider Hermitian forms. If σ is an automorphism of $\mathrm{GF}(q)$ of order 2, then q is a square, say $q=r^2$, and $\alpha^\sigma=\alpha^r$. We need the fact that every element of $\mathrm{Fix}(\sigma)=\mathrm{GF}(r)$ has the form $\alpha\alpha^\sigma$ (see Exercise 1 of Section 6.2).

In one variable, we have $f(x)=\mu x x^\sigma$ for some non-zero $\mu \in \mathrm{Fix}(\sigma)$; writing $\mu=\alpha\alpha^\sigma$ and replacing x by αx, we can assume that $\mu=1$.

In two variables, we can similarly take the form to be $xx^\sigma+yy^\sigma$. Now $-1 \in \mathrm{Fix}(\sigma)$, so $-1=\lambda\lambda^\sigma$; then the form vanishes at $(1,\lambda)$. It follows that there is no anisotropic form in any larger number of variables either. ■

Exercises

1. Let b be a σ-Hermitian form on a vector space V over F, where σ is not the identity. Set $f(\mathbf{v})=b(\mathbf{v},\mathbf{v})$. Let $E=\mathrm{Fix}(\sigma)$, and let V' be V regarded as an E-vector space by restricting scalars. Prove that f is a quadratic form on V', which polarises to the bilinear form $\mathrm{Tr}(b)$ defined by $\mathrm{Tr}(b)(\mathbf{v},\mathbf{w})=b(\mathbf{v},\mathbf{w})+b(\mathbf{v},\mathbf{w})^\sigma$. Show further that $\mathrm{Tr}(b)$ is non-degenerate if and only if b is.

2. Prove that there is, up to equivalence, a unique non-degenerate alternating bilinear form on a vector space of countably infinite dimension (a direct sum of countably many isotropic lines).

3. Let F be a finite field of characteristic 2.

(a) Prove that every element of F has a unique square root.

(b) By considering the bilinear form obtained by polarisation, prove that a non-singular form in 2 or 3 variables over F is equivalent to $\alpha x^2 + xy + \beta y^2$ or $\alpha x^2 + xy + \beta y^2 + \gamma z^2$ respectively. Prove that forms of the first shape (with $\alpha, \beta \neq 0$) are all equivalent, while those of the second shape cannot be anisotropic.

6.4 Classical polar spaces

Polar spaces describe the geometry of vector spaces carrying a reflexive sesquilinear form or a quadratic form in much the same way as projective spaces describe the geometry of vector spaces. We now embark on the study of these geometries; the three preceding sections contain the prerequisite algebra.

First, some terminology. The polar spaces associated with the three types of forms (alternating bilinear, Hermitian, and quadratic) are referred to by the same names as the groups associated with them: *symplectic*, *unitary*, and *orthogonal* respectively. Of what do these spaces consist?

Let V be a vector space carrying a form of one of our three types. Recall that as well as a sesquilinear form b in two variables, we have a form f in one variable — either f is defined by $f(\mathbf{v}) = b(\mathbf{v},\mathbf{v})$, or b is obtained by polarising f — and we make use of both forms. A subspace of V on which b vanishes identically is called a *totally isotropic subspace* (or *t.i. subspace*), while a subspace on which f vanishes identically is called a *totally singular subspace* (or *t.s. subspace*). Every t.s. subspace is t.i., but the converse is false. In the case of alternating forms, every subspace is t.s.! I frequently use the expression *t.i. or t.s. subspace*, to mean a t.i. subspace (in the symplectic or unitary case) or a t.s. subspace (in the orthogonal case).

The *classical polar space* (or simply the *polar space*) associated with a vector space carrying a form is the geometry whose flats are the t.i. or t.s. subspaces (in the above sense). (Concerning the terminology: the term "polar space" is normally reserved for a geometry satisfying the axioms of Tits, which we will meet shortly. But every classical polar space is a polar space, so the terminology here should cause no confusion.) Note that, if the form is anisotropic, then the only member of the polar space is the zero subspace. The *polar rank* of a classical polar space is

the largest vector space rank of any t.i. or t.s. subspace; it is zero if and only if the form is anisotropic. Where there is no confusion, polar rank will be called simply *rank*. (We will soon see that there is no conflict with our earlier definition of polar rank as the number of hyperbolic lines in the decomposition of the space.) We use the terms *point, line, plane,* etc., just as for projective spaces.

We now proceed to derive some properties of polar spaces. Let G be a classical polar space of polar rank r.

First, we identify the two definitions of polar space rank. We use the expression for V as the direct sum of r hyperbolic lines and an anisotropic subspace given by Theorem 6.7. Any t.i. or t.s. subspace meets each hyperbolic line in at most a point, and meets the anisotropic germ in the zero space; so its rank is at most r. But the span of r t.i. or t.s. points, one chosen from each hyperbolic line, is a t.i. or t.s. subspace of rank r.

(P1) Any flat, together with the flats it contains, is a projective space of dimension at most $r-1$.

This is clear since a subspace of a t.i. or t.s. subspace is itself t.i. or t.s. The next property is also clear.

(P2) The intersection of any family of flats is a flat.

(P3) If U is a flat of dimension $r-1$ and p a point not in U, then the union of the lines joining p to points of U is a flat W of dimension $r-1$; and $U \cap W$ is a hyperplane in both U and W.

Proof Let $p = \langle \mathbf{w} \rangle$. The function $\mathbf{v} \mapsto b(\mathbf{v}, \mathbf{w})$ on the vector space U is linear; let K be its kernel, a hyperplane in U. Then the line (of the projective space) joining p to a point $q \in U$ is t.i. or t.s. if and only if $q \in K$; and the union of all such t.i. or t.s. lines is a t.i. or t.s. space $W = \langle K, \mathbf{v} \rangle$, such that $W \cap U = K$, as required.

(P4) There exist two disjoint flats of dimension $r-1$.

Proof Use the hyperbolic-anisotropic decomposition again. If $L_1, \ldots, L_r$ are the hyperbolic lines, and $\mathbf{v}_i, \mathbf{w}_i$ are the distinguished spanning vectors in L_i, then the required flats are $\langle \mathbf{v}_1, \ldots, \mathbf{v}_r \rangle$ and $\langle \mathbf{w}_1, \ldots, \mathbf{w}_r \rangle$.

Next, we specialise to the case $r = 2$. (A polar space of rank 1 is just an unstructured collection of points.) A polar space of rank 2 consists of points and lines, and has the following properties. (The first two are immediate consequences of (P2) and (P3) respectively.)

(Q1) Two points lie on at most one line.

(Q2) If L is a line, and p a point not on L, then there is a unique point of L collinear with p.

(Q3) No point is collinear with all others.

For, by (P4), there exist disjoint lines; and, given any point p, at least one of these lines does not contain p, and p fails to be collinear with some point of this line.

A geometry satisfying (Q1), (Q2) and (Q3) is called a *generalised quadrangle*. Such geometries play much the same rôle in the theory of polar spaces as projective planes do in the theory of projective spaces. We will return to them later.

Note that (Q1) holds in a polar space of arbitrary rank.

Another property of polar spaces, which is proved by almost the same argument as (P3), is the following extension of (Q2):

(BS) If L is a line, and p a point not on L, then p is collinear with one or all points of L.

In a polar space G, for any set S of points, we let $S^\perp$ denote the set of points which are perpendicular to (that is, collinear with) every point of S. It follows from (BS) that, for any set S, the set $S^\perp$ is a (linear) subspace of G (that is, if two points of $S^\perp$ are collinear, then the line joining them lies wholly in $S^\perp$). Moreover, for any point x, $x^\perp$ is a hyperplane of G (that is, a subspace which meets every line).

Polar spaces have good inductive properties. Let G be a classical polar space. There are two natural ways of producing a "smaller" polar space from G:

(a) Take a point x of G, and consider the quotient space $x^\perp/x$, the space whose points, lines, ... are the lines, planes, ... of G containing x.

(b) Take two non-perpendicular points x and y, and consider $\{x,y\}^\perp$.

In each case, the space constructed is a classical polar space, having the same germ as G but with polar rank one less than that of G. (Note that, in (b), the span of x and y in the vector space is a hyperbolic line.) There are more general versions. For example, if S is a flat of dimension $d-1$, then $S^\perp/S$ is a polar space

of rank $r-d$ with the same germ as G. We will see below and in the next section how this inductive process can be used to obtain information about polar spaces.

We investigate just one type in more detail, the so-called *hyperbolic quadric* or *hyperbolic orthogonal space*, the orthogonal space which is a direct sum of hyperbolic lines (that is, having germ 0). The quadratic form defining this space can be taken to be $x_1x_2 + x_3x_4 + \ldots + x_{2r-1}x_{2r}$.

Theorem 6.11 *The maximal flats of a hyperbolic quadric fall into two classes, with the properties that the intersection of two maximal flats has even codimension in each if and only if they belong to the same class.*

Proof First, note that the result holds when $r = 1$, since then the quadratic form is x_1x_2 and there are just two singular points, $\langle(1,0)\rangle$ and $\langle(0,1)\rangle$. By the inductive principle, it follows that any flat of dimension $r-2$ is contained in exactly two maximal flats.

We take the $(r-1)$-flats and $(r-2)$-flats as the vertices and edges of a graph Γ, that is, we join two $(r-1)$-flats if their intersection is an $(r-2)$-flat. The theorem will follow if we show that Γ is connected and bipartite, and that the distance between two vertices of Γ is the codimension of their intersection. Clearly the codimension of the intersection increases by at most one with every step in the graph, so it is at most equal to the distance. We prove equality by induction.

Let U be a $(r-1)$-flat and K a $(r-2)$-flat. We claim that the two $(r-1)$-spaces W_1, W_2 containing K have different distances from U. Factoring out the t.s. subspace $U \cap K$ and using induction, we may assume that $U \cap K = \emptyset$. Then $U \cap K^\perp$ is a point p, which lies in one but not the other of W_1, W_2; say $p \in W_1$. By induction, the distance from U to W_1 is $r-1$; so the distance from U to W_2 is at most r, hence equal to r by the remark in the preceding paragraph.

This establishes the claim about the distance. The fact that Γ is bipartite also follows, since in any non-bipartite graph there exists an edge both of whose vertices have the same distance from some third vertex, and the argument given shows that this doesn't happen in Γ. ∎

In particular, the rank 2 hyperbolic quadric consists of two families of lines forming a *grid*, as shown in Fig. 6.1. This is the so-called "ruled quadric", familiar from models such as wastepaper baskets.

Exercises

1. Prove (BS).

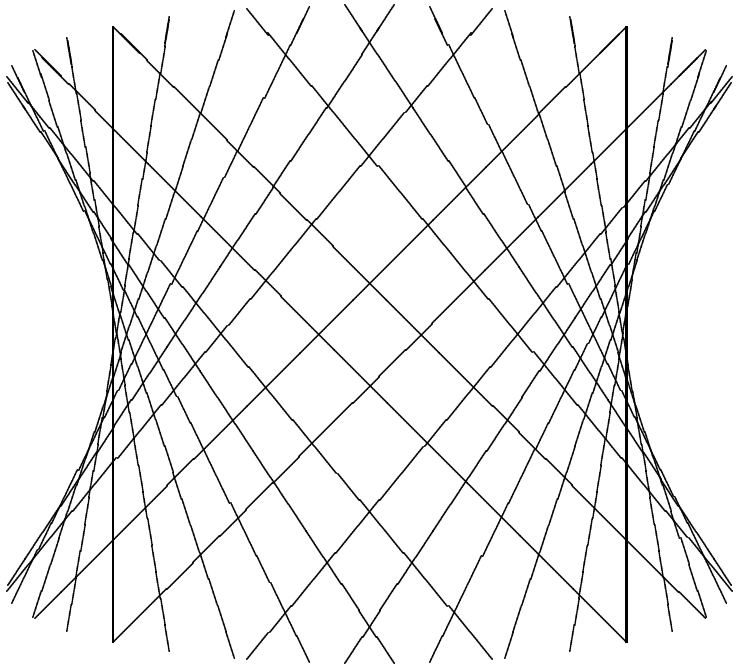


Figure 6.1: A grid

2. Prove the assertions above about $x^\perp/x$ and $\{x,y\}^\perp$.

3. Show that Theorem 6.11 can be proved using only properties (P1)–(P4) of polar spaces together with the fact that an $(r-1)$-flat lies in exactly two maximal flats.

6.5 Finite polar spaces

The classification of finite classical polar spaces was achieved by Theorem 6.7. We subdivide these spaces into six families according to their germ, viz., one symplectic, two unitary, and three orthogonal. (Forms which differ only by a scalar factor obviously define the same polar space.) The following table gives some information about them. In the table, r denotes the polar space rank, n the vector space rank. The significance of the parameter ε will emerge shortly. This number, depending only on the germ, carries numerical information about all spaces in the family. Note that, in the unitary case, the order of the finite field

must be a square.

Type	n	ε
Symplectic	$2r$	0
Unitary	$2r$	$-\frac{1}{2}$
Unitary	$2r+1$	$\frac{1}{2}$
Orthogonal	$2r$	-1
Orthogonal	$2r+1$	0
Orthogonal	$2r+2$	1

Table 6.1: Finite classical polar spaces

Theorem 6.12 *The number of points in a finite polar space of rank 1 is $q^{1+\varepsilon}+1$, where ε is given in Table 6.1.*

Proof Let V be a vector space carrying a form of rank 1 over $\mathrm{GF}(q)$. Then V is the orthogonal direct sum of a hyperbolic line L and an anisotropic germ U of dimension k (say). Let n_k be the number of points.

Suppose that $k > 0$. If p is a point of the polar space, then p lies on the hyperplane $p^\perp$; any other hyperplane containing p is non-degenerate with polar rank 1 and having germ of dimension $k-1$. Consider a parallel class of hyperplanes in the affine space whose hyperplane at infinity is $p^\perp$. Each such hyperplane contains $n_{k-1}-1$ points, and the hyperplane at infinity contains just one, viz., p. So we have

$$n_k - 1 = q(n_{k-1}-1),$$

from which it follows that $n_k = 1+(n_0-1)q^k$. So it is enough to prove the result for the case $k = 0$, that is, for a hyperbolic line.

In the symplectic case, each of the $q+1$ projective points on a line is isotropic.

Consider the unitary case. We can take the form to be

$$b((x_1,y_1),(x_2,y_2)) = x_1\overline{y_2} + y_1\overline{x_2},$$

where $\overline{x} = x^\sigma = x^r$, $r^2 = q$. So the isotropic points satisfy $x\overline{y} + y\overline{x} = 0$, that is, $\mathrm{Tr}(x\overline{y}) = 0$. How many pairs (x,y) satisfy this? If $y = 0$, then x is arbitrary. If $y \neq 0$, then a fixed multiple of x is in the kernel of the trace map, a set of size $q^{1/2}$ (since Tr is $\mathrm{GF}(q^{1/2})$-linear). So there are

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

vectors, i.e., $q^{1/2}+1$ projective points.

Finally, consider the orthogonal case. The quadratic form is equivalent to xy, and has two singular points, $\langle(1,0)\rangle$ and $\langle(1,0)\rangle$. ■

Theorem 6.13 *In a finite polar space of rank r, there are $(q^r-1)(q^{r+\varepsilon}+1)/(q-1)$ points, of which $q^{2r-1+\varepsilon}$ are not perpendicular to a given point.*

Proof We let $F(r)$ be the number of points, and $G(r)$ the number not perpendicular to a given point. (We do not assume that $G(r)$ is constant; this constancy follows from the induction that proves the theorem.) We use the two inductive principles described at the end of the last section.

Step 1 $G(r)=q^2G(r-1)$.

Take a point x, and count pairs (y,z), where $y\in x^\perp$, $z\notin x^\perp$, and $z\in y^\perp$. Choosing z first, there are $G(r)$ choices; then $\langle x,z\rangle$ is a hyperbolic line, and y is a point in $\langle x,z\rangle^\perp$, so there are $F(r-1)$ choices for y. On the other hand, choosing y first, the lines through y are the points of the rank $r-1$ polar space $x^\perp/x$, and so there are $F(r-1)$ of them, with q points different from x on each, giving $qF(r-1)$ choices for y; then $\langle x,y\rangle$ and $\langle y,z\rangle$ are non-perpendicular lines in $y^\perp$, i.e., points of $y^\perp/y$, so there are $G(r-1)$ choices for $\langle y,z\rangle$, and so $qG(r-1)$ choices for y. thus

$$G(r)\cdot F(r-1)=qF(r-1)\cdot qG(r-1),$$

from which the result follows.

Since $G(1)=q^{1+\varepsilon}$, it follows immediately that $G(r)=q^{2r-1+\varepsilon}$, as required.

Step 2 $F(r)=1+qF(r-1)+G(r)$.

For this, simply observe (as above) that points perpendicular to x lie on lines of $x^\perp/x$.

Now it is just a matter of calculation that the function $(q^r-1)(q^{r+\varepsilon}+1)/(q-1)$ satisfies the recurrence of Step 2 and correctly reduces to $q^{1+\varepsilon}+1$ when $r=1$. ■

Theorem 6.14 *The number of maximal flats in a finite polar space of rank r is*

$$\prod_{i=1}^{r}(1+q^{i+\varepsilon}).$$

Proof Let $H(r)$ be this number. Count pairs (x,U), where U is a maximal flat and $x \in U$. We find that

$$F(r) \cdot H(r-1) = H(r) \cdot (q^r - 1)/(q-1),$$

so

$$H(r) = (1 + q^{r+\varepsilon})H(r-1).$$

Now the result is immediate. ■

It should now be clear that any reasonable counting question about finite polar spaces can be answered in terms of q, r, ε.

7

Axioms for polar spaces

The axiomatisation of polar spaces was begun by Veldkamp, completed by Tits, and simplified by Buekenhout, Shult, Hanssens, and others. In this chapter, the analogue of Chapter 3, these results are discussed, and proofs given in some cases as illustrations. We begin with a discussion of generalised quadrangles, which play a similar rôle here to that of projective planes in the theory of projective spaces.

7.1 Generalised quadrangles

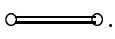
We saw the definition of a generalised quadrangle in Section 6.4: it is a rank 2 geometry satisfying the conditions

(Q1) two points lie on at most one line;

(Q2) if the point p is not on the line L, then p is collinear with exactly one point of L;

(Q3) no point is collinear with all others.

For later use, we represent generalised quadrangles by a diagram with a double arc, thus:



The axioms (Q1)–(Q3) are self-dual; so the dual of a generalised quadrangle is also a generalised quadrangle.

Two simple classes of examples are provided by the *complete bipartite graphs*, whose points fall into two disjoint sets (with at least two points in each, and whose

lines consist of all pairs of points containing one from each set), and their duals, the *grids*, some of which we met in Section 6.4. Any generalised quadrangle in which lines have just two points is a complete bipartite graph, and dually (Exercise 2). We note that any line contains at least two points, and dually: if L were a singleton line $\{p\}$, then every other point would be collinear with p (by (Q2)), contradicting (Q3).

Apart from complete bipartite graphs and grids, all generalised quadrangles have orders:

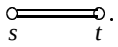
Theorem 7.1 *Let G be a generalised quadrangle in which there is a line with at least three points and a point on at least three lines. Then the number of points on a line, and the number of lines through a point, are constants.*

Proof First observe that, if lines L_1 and L_2 are disjoint, then they have the same cardinality; for collinearity sets up a bijection between the points on L_1 and those on L_2.

Now suppose that L_1 and L_2 intersect. Let p be a point on neither of these lines. Then one line through p meets L_1, and one meets L_2, so there is a line L_3 disjoint from both L_1 and L_2. It follows that L_1 and L_2 both have the same cardinality as L_3.

The other assertion is proved dually. ■

This proof works in both the finite and the infinite case. If G is finite, we let s and t be the orders of G; that is, any line has $s+1$ points and any point lies on $t+1$ lines, so that the diagram is



For the classical polar spaces over $\mathrm{GF}(q)$, we have $s = q$ and $t = q^{1+\varepsilon}$, where ε is given in Table 6.5.1.

From now on, "generalised quadrangle" will be abbreviated to GQ.

The next result summarises some properties of finite GQs.

Theorem 7.2 *Let G be a finite GQ with orders s,t.*

(a) G has $(s+1)(st+1)$ points and $(t+1)(st+1)$ lines.

(b) $s+t$ divides $st(s+1)(t+1)$;

(c) if $s>1$, then $t \le s^2$;

(d) if $t > 1$, then $s \leq t^2$.

Proof (a) is proved by elementary counting, like that in Section 6.5. (b) is shown by an argument involving eigenvalues of matrices, in the spirit of the proof of the Friendship Theorem outlined in Exercise 2.2.4. (c) is proved by elementary counting (see Exercise 3), and (d) is dual to (c). ■

In particular, if $s = 2$, then $t \leq 4$; and the case $t = 3$ is excluded by (b) above. So $t = 1, 2$ or 4. These three values are realised by the three orthogonal rank 2 polar spaces over $\mathrm{GF}(2)$. We will see, as a special case of a later result, that these are the only GQs with $s = 2$. However, this result is sufficiently interesting to be worth another proof which generalises it in a different direction.

Theorem 7.3 *Let G be a GQ with orders $s = 2$ and t. Then $t = 1, 2$ or 4; and there is a unique geometry for each value of t.*

Note the generalisation: t is not assumed to be finite!

Proof Take a point and call it ∞; let $\{L_i : i \in I\}$ be the set of lines containing ∞. Number the points other than ∞ on L_i as p_{i0} and p_{i1}. Now, for any point q not collinear with p, there is a function $f_q : I \to \{0, 1\}$ defined by the rule that the unique point of L_i collinear with q is $p_{i f_q(i)}$. We use the function f_q as a label for q. Let X be the set of points not collinear with ∞. We consider the possible relationships of points in X. Write $q \sim r$ if q and r are collinear.

1. If $q, r \in X$ satisfy $q \sim r$, then f_q and f_r agree in just one position, viz., the unique index i for which the line L_i through ∞ meets the line qr.

2. If q, r are not collinear but some point of X is collinear with both, then f_q and f_r agree in all but two positions; for all but two values are changed twice, the remaining two being changed just once.

3. Otherwise, $f_q = f_r$; for all the common neighbours of q and r are adjacent to ∞.

Note too that, for any $i \in I$ and $q \in X$, there is a point $r \sim q$ for which f_q and f_r agree only in i, viz., the last point of the line through q meeting L_i.

Now suppose (as we may) that $|I| > 2$, and choose distinct $i, j, k \in I$. Given $q \in X$, choose $r, s, t \in X$ such that f_q and f_r agree only in i, f_r and f_s only in j, and f_s and f_t only on k. Then clearly f_q and f_t agree in precisely the three points i, j, k, since these values are changed twice and all others three times. By the case analysis, it follows that $|I| = 3$ or $|I| - 2 = 3$, as required.

The uniqueness also follows from this analysis, with a little more work: we know enough about the structure of X that the entire geometry can be reconstructed. ■

Problem. Can there exist a GQ with s finite ($s > 1$) and t infinite?

The proof above shows that there is no such GQ with $s = 2$. It is also known that there is no GQ with $s = 3$ or $s = 4$ and t infinite, though the proofs are much harder. (This is due to Kantor and Brouwer for $s = 3$, and Cherlin for $s = 4$.) Beyond this, nothing is known, though Cherlin's argument could in principle be extended to larger values of s.

The GQs with $s = 2$ and $t = 1, 2$ have simple descriptions. For $s = 1$ we have the 3×3 grid. For $t = 2$, take the points to be all the 2-element subsets of a set of cardinality 6, and the lines to be all partitions of the 6-set into three disjoint 2-subsets. The GQ with order $(2,4)$ is a little harder to describe. The implicit construction in Theorem 7.3 is one of the simplest — the functions f_q are all those which take the value 1 an even number of times, each such function representing a unique point. This GQ also arises in classical algebraic geometry, as the Schläfli configuration of 27 lines in a general cubic surface, lying three at a time in 45 planes.

In the classical polar spaces, the orders s and t are both powers of the same prime. There are examples where this is not the case — see Exercise 5.

Exercises

1. Prove that the dual of a GQ is a GQ.
2. Prove that a GQ with two points on any line is a complete bipartite graph.
3. Let G be a finite GQ with orders s, t, where $s > 1$. Let p_1 and p_2 be non-adjacent points, and let x_n be the number of points p_3 adjacent to neither p_1 nor p_2 for which there are exactly n common neighbours of p_1, p_2 and p_3. Show that

$$\begin{aligned} \sum x_n &= s^2t - st - s + t, \\ \sum n x_n &= s(t+1)(t-1), \\ \sum n(n-1)x_n &= (t+1)t(t-1). \end{aligned}$$

Hence prove that $t \leq s^2$, with equality if and only if any three pairwise non-collinear points have exactly $s+1$ common neighbours.

4. In this exercise, we use the terminology of coding theory (as in Section 3.2). Consider the space V of words of length 6 with even weight. This is a vector space of rank 5 over $\mathrm{GF}(2)$. The "standard inner product" on V is a bilinear form which is alternating (by the even weight condition); its radical $V^\perp$ is spanned by the unique word of weight 6. Thus, $V/V^\perp$ is a vector space of rank 4 carrying a non-degenerate alternating bilinear form. The 15 non-zero vectors of this space

are cosets of $V^\perp$ containing a word of weight 2 and the complementary word of weight 4, and so can be identified with the 2-subsets of a 6-set. Extend this identification to an isomorphism between the combinatorial description of the GQ with orders (2, 2) and the rank 2 symplectic polar space over $\mathrm{GF}(2)$.

5. Let q be an even prime power, and let C be a hyperoval in $\Pi = \mathrm{PG}(2,q)$, a set of $q+2$ points meeting every line in 0 or 2 points (see Section 4.3). Now take Π to be the hyperplane at infinity of $\mathrm{AG}(3,q)$. Let G be the geometry whose points are all the points of $\mathrm{AG}(3,q)$, and whose lines are all the lines of $\mathrm{AG}(3,q)$ which meet Π in a point of C. Prove that G is a GQ with orders $(q-1,q+1)$.

6. Construct "free" GQs.

7.2 Diagrams for polar spaces

The inductive properties of polar spaces are exactly what is needed to show that they are diagram geometries.

Proposition 7.4 *A classical polar space of rank n belongs to the diagram*

with n nodes.

Proof Given a variety U of rank d, the varieties contained in it form a projective space of dimension $d-1$, while the varieties containing it are those of the polar space $U^\perp/U$ of rank $n-d$; moreover, any variety contained in U is incident with any variety containing U. Since a rank 2 polar space is a generalised quadrangle, it follows by induction that residues of varieties are correctly described by the diagram. ■

This diagram is commonly referred to as C_n.

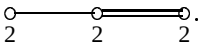
By analogy with Section 5.2, it might be thought that any geometry with diagram C_n for $n \geq 3$ is a classical polar space. This is false for several reasons, which we will see at various points. But first, here is one example of a geometry with diagram C_3 which is nothing like a polar space, even though it is highly symmetrical. This geometry was discovered by Arnold Neumaier, and is referred to as *Neumaier's geometry* or the *A_7-geometry*.

Let X be a set of seven points. The structure of a projective plane $\mathrm{PG}(2,2)$ can be imposed on X in 30 different ways — this number is the index of $\mathrm{PGL}(3,2)$

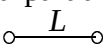
in the symmetric group S_7. Since $\mathrm{PGL}(3,2)$ contains no odd permutations, it is contained in the alternating group A_7 with index 15, and so the 30 planes fall into two orbits of length 15 under A_7. Now we take the points, lines, and planes of the geometry to be respectively the elements of X, the 3-element subsets of X, and one orbit of A_7 on $\mathrm{PG}(2,2)$s. Incidence between points and lines, or between lines and planes, is defined by membership; and every point is incident with every plane.

It is clear that the residue of a plane is a projective plane $\mathrm{PG}(2,2)$, while the residue of a line is a digon. Consider the residue of a point x. The lines incident with x can be identified with the 2-element subsets of the 6-element set $Y = X \setminus \{x\}$. Given a plane, its three lines containing x partition Y into three 2-sets. It is easy to check that, given such a triple of lines, there are just two ways to draw the remaining four lines to complete $\mathrm{PG}(2,2)$, and that these two are related by an odd permutation of X. So our chosen orbit of planes has exactly one member inducing the given partition of Y, and the planes incident with x can be identified with all the partitions of Y into three 2-sets. As we saw in Section 7.1, this incidence structure is a generalised quadrangle with order $2,2$.

We conclude that the geometry has the diagram



.

This example shows that, even in a geometry with such a simple diagram, a variety is not necessarily determined by its point-shadow (all planes have the same point-shadow!); the intersection of point-shadows of varieties need not be the point-shadow of a variety, and the points and lines need not form a partial linear space. So the special properties of linear diagrams with all strokes 

do not extend. However, classical polar spaces do have these nice properties.

A C_3-geometry in which every point and every plane are incident is called *flat*. Neumaier's geometry is the only known finite example of such a geometry. Some infinite examples were constructed by Sarah Rees; we now describe these. First, a re-interpretation of Neumaier's geometry.

Consider the rank 6 vector space V of all binary words of length 7 having even weight. On V, we can define a quadratic form by the rule

$$f(\mathbf{v}) = \tfrac{1}{2}\,\mathrm{wt}(\mathbf{v}) \pmod 2.$$

The bilinear form obtained by polarising f is just the usual dot product, since

$$\mathrm{wt}(\mathbf{v}+\mathbf{w}) = \mathrm{wt}(\mathbf{v}) + \mathrm{wt}(\mathbf{w}) - 2\mathbf{v}.\mathbf{w}\,.$$

It follows that f is non-singular: the only vector orthogonal to V is the all-1 word, which is not in V. Now the points of X, which index the coordinates, are in one-one correspondence with the seven words of weight 6, which are non-singular vectors. The lines correspond to the vectors of weight 4, which comprise all the singular vectors.

We saw in Section 6.4 that the planes on the quadric fall into two families, such that two planes of the same family meet in a subspace of even codimension (necessarily a point), while planes of different families meet in a subspace of odd codimension (the empty set or a line). Now a plane on the quadric contains seven non-zero singular vectors (of weight 4), any two of which are orthogonal, and so meet in an even number of points, necessarily 2. The complements of these 4-sets form seven 3-sets, any two meeting in one point, so forming a projective plane $\mathrm{PG}(2,2)$. It is readily checked that the two classes of planes correspond exactly to the two orbits of A_7 we described earlier. So the points, lines and planes of Neumaier's geometry can be identified with a special set of seven non-singular points, the singular points, and one family of planes on the quadric. Incidence between the non-singular and the singular points is defined by orthogonality.

Now we reverse the procedure. We start with a hyperbolic quadric Q in $\mathrm{PG}(5,F)$, that is, a quadric of rank 3 with germ zero. A set S of non-singular points is called an *exterior set* if it has the property that, given any line L of Q, a unique point of S is orthogonal to L. Now consider the geometry G whose POINTS, LINES and PLANES are the points of S, the points of Q, and one family of planes on Q; incidence between POINTS and LINES is defined by orthogonality, that between LINES and PLANES is incidence in the polar space, and every POINT is incident with every PLANE.

Such a geometry belongs to the diagram C_3. For the residue of a PLANE Π is a projective plane, naturally the dual of Π. (The correspondence between points of S and lines of Π is bijective; for, given $x \in S$, $x^\perp$ cannot contain Π, since a polar space in $\mathrm{PG}(4,q)$ cannot have rank 3, and so it meets Π in a line.) The residue of a POINT x is the polar space $x^\perp$, which as we've seen is rank 2, and so a GQ. And of course the POINTS and PLANES incident with a LINE form a digon.

Thas showed that no further finite examples can be constructed in this way:

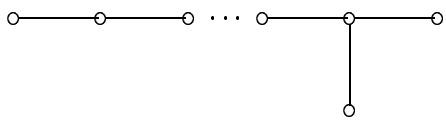
Theorem 7.5 *There is no exterior set for the hyperbolic quadric in* $\mathrm{PG}(5,q)$ *for* $q > 2$.

However, Rees (who first described this construction) observed that there are infinite examples. Consider the case where $F = \mathbb{R}$; let the form be $x_1x_2 + x_3x_4 +$

x_5x_6. Now the space of rank 3 spanned by $(1,1,0,0,0,0)$, $(0,0,1,1,0,0)$ and $(0,0,0,0,1,1)$ is positive definite, and so is disjoint from the quadric; the points spanned by vectors in this space form an exterior set.

Now we turn to hyperbolic quadrics in general. As we saw in Section 6.4, the maximal t.s. subspaces on such a quadric Q of rank n can be partitioned into two families, so that a flat of dimension $n-2$ lies in a unique member of each family. We construct a new geometry by letting these flats be varieties of different types. Now there is no need to retain the flats of dimension $n-2$, since such a flat is the intersection of the two maximal flats containing it.

Theorem 7.6 *Let Q be a hyperbolic quadric of rank $n \geq 3$. Let G be the geometry whose flats are the t.s. subspaces of dimension different from $n-2$, where the two families of flats of dimension $n-1$ are assigned different types. Incidence between flats, at least one of which has dimension less than $n-1$, is as usual; while $(n-1)$-flats of different types are incident if they intersect in an $(n-2)$-flat. Then the geometry has diagram*



(n nodes).

Proof We need only check the residue of a flat of dimension $n-3$: the rest follows by induction, as in Proposition 7.4. Such a flat cannot be the intersection of two $(n-1)$-flats of different types; so any two such flats of different types containing it are incident. ■

This diagram is denoted by D_n. The result holds also for $n=2$, provided that we interpret D_2 as two unconnected nodes — the quadric has two families of lines, each line of one family meeting each line of the other.

Exercises

1. Prove that the line joining two points of an exterior set to the quadric Q is disjoint from Q.

2. Prove that an exterior set to a quadric in $\mathrm{PG}(5,q)$ must have q^2+q+1 points.

3. Show that the plane constructed in Rees' example is an exterior set.

7.3 Tits and Buekenhout–Shult

We now begin working towards the axiomatisation of polar spaces. This major result of Tits (building on earlier work of Veldkamp) will not be proved completely here, but the next four sections should give some impression of how the proof works.

Tits' theorem characterises a class of spaces which almost coincides with the classical polar spaces of rank at least 3. There are a few additional examples of rank 3, some of which will be described later. I will use the term *abstract polar space* for a geometry satisfying the axioms. In fact, Tits' axioms describe all subspaces of arbitrary dimension; an alternative axiom system, proposed by Buekenhout and Shult, involves only points and lines (in the spirit of the Veblen–Young axioms for projective spaces). In this section, I show the equivalence of these axiom systems.

Temporarily, then, an *abstract polar space of type T* is a geometry satisfying the conditions (P1)–(P4) of Section 6.4, repeated here for convenience.

(P1) Any flat, together with the flats it contains, is a projective space of dimension at most $r-1$.

(P2) The intersection of any family of flats is a flat.

(P3) If U is a flat of dimension $r-1$ and p a point not in U, then the union of the lines joining p to points of U is a flat W of dimension $r-1$; and $U \cap W$ is a hyperplane in both U and W.

(P4) There exist two disjoint flats of dimension $r-1$.

An *abstract polar space of type BS* is a geometry of points and lines satisfying the following conditions. In these axioms, a *subspace* is a set S of points with the property that if a line L contains two points of S, then $L \subseteq S$; a *singular subspace* is a subspace, any two of whose points are collinear.

(BS1) Any line contains at least three points.

(BS2) No point is collinear with all others.

(BS3) Any chain of singular subspaces is of finite length.

(BS4) If the point p is not on the line L, then p is collinear with one or all points of L.

(Note that (BS4) is our earlier (BS), and is the key condition here.)

Theorem 7.7 *(a) The points and lines of an abstract polar space of type T form an abstract polar space of type BS.*

(b) The singular subspaces of an abstract polar space of type BS form an abstract polar space of type T.

Proof (a) It is an easy deduction from (P1)–(P4) that any subspace is contained in a subspace of dimension $n-1$. For let U be a subspace, and W a subspace of dimension $n-1$ for which $U \cap W$ has dimension as large as possible; if $p \in U \setminus W$, then (P3) gives a subspace of dimension $n-1$ containing p and $U \cap W$, contradicting maximality.

Now, if L is a line and p a point not on L, let W be a subspace of dimension $n-1$ containing L. If $p \in W$, then p is collinear with every point of L; otherwise, the neighbours of p in W form a hyperplane, meeting L in one or all of its points.

Thus, (BS4) holds. The other conditions are clear.

(b) Now let G be an abstract polar space of type BS. Call two points *adjacent* if they are collinear; this gives the point set a graph structure. Every maximal clique in the graph is a subspace. For let S be a maximal clique, and $p, q \in S$; let L be a line containing p and q. Any point of $S \setminus L$ is collinear with p and q, and so with every point of L; thus $S \cup L$ is a clique, and by maximality, $L \subseteq S$.

If $p \notin S$ (where S is a maximal clique), then the set of neighbours of p in S is a hyperplane. Every point $q \in S$ lies outside such a hyperplane; for, by (BS2), there is a point p not adjacent to q. As we saw in Section 3.1, if every line has size 3, then this implies that S is a projective space; but this deduction cannot be made in general. However, in the present situation, Buekenhout and Shult are able to show that S is indeed a projective space. (In particular, this implies that two points lie on at most one line. For the union of two lines through two common points is a clique by (BS4), and so would be contained in a maximal clique. However, Buekenhout and Shult have to show that two points lie on at most one line before they know that the subspaces are projective spaces; the proof is surprisingly tricky.)

Any singular subspace lies in some maximal clique, and so is itself a projective space. Thus (P1) holds; and the remaining axioms can now be verified. ■

We will now simplify the terminology by using the term "abstract polar space" equally for either type.

The induction principles we used in classical polar spaces work in almost the same way in abstract polar spaces.

Proposition 7.8 *Let U be a $(d-1)$-dimensional subspace of an abstract polar space of rank n. Then the subspaces containing U form an abstract polar space of rank $n-d$.* ■

Exercise

1. Show directly that, in an abstract polar space of type BS having three points on any line, any two points lie on at most one line, and singular subspaces are projective.

7.4 Recognising hyperbolic quadrics

There are two special cases where the proof of the characterisation of polar spaces is substantially easier, namely, hyperbolic quadrics and quadrics over $\mathrm{GF}(2)$; they will be treated in this section and the next.

In the case of a hyperbolic quadric, we bypass the need to reconstruct the quadric by simply showing that there is a unique example of each rank over any field. First, we observe that the partition of the maximal subspaces into two types follows directly from the axioms; properties of the actual model are not required. We begin with a general result on abstract polar spaces.

An abstract polar space G can be regarded as a point-line geometry, as we've seen. Sometimes it is useful to consider a "dual" situation, defining a geometry G^* whose POINTs are the maximal subspaces of G and whose LINEs are the next-to-maximal subspaces, incidence being reversed inclusion. We call this geometry a *dual polar space*. In a dual polar space, we define the distance between two POINTs to be the number of LINEs on a shortest path joining them.

Proposition 7.9 *Let G^* be a dual polar space.*

(a) The distance between two POINTs is the codimension of their intersection.

(b) Given a POINT p and a LINE L, there is a unique POINT of L nearest to p.

Proof Let U_1, U_2 be maximal subspaces. By the inductive principle (Proposition 7.8), we may assume that $U_1 \cap U_2 = \emptyset$. (It is clear that any path from U_1 to U_2, in which not all terms contain $U_1 \cap U_2$, must have length strictly greater than the codimension of $U_1 \cap U_2$; so, once the result is proved in the quotient, no such path can be minimal.)

Now each point of U_1 is collinear (in G) with a hyperplane in U_2, and *vice versa*; so, given any hyperplane H in U_2, there is a unique point of U_1 adjacent to H, and hence (by (P3)) a unique maximal subspace containing H and meeting U_1. The result follows.

(b) Let U be a maximal subspace and W a subspace of rank one less than maximal. As before, we may assume that $U \cap W = \emptyset$. Now there is a unique point $p \in U$ collinear with all points of W. Then $\langle W, p\rangle$ is the uniue POINT on the LINE W nearest to the POINT U. ∎

Proposition 7.10 *Let G be an abstract polar space of rank n, in which any $(n-2)$-dimensional subspace is contained in exactly two maximal subspaces. Then the maximal subspaces fall into two families, the intersection of two subspaces having even codimension in each if and only if the subspaces belong to the same family.*

Proof The associated dual polar space is a graph. By Proposition 7.9(b), the graph is bipartite, since if an odd circuit exists, then there is one of minimal length, and both vertices on any edge are then equidistant from the opposite vertex in the cycle. ∎

Now, in any abstract polar space of rank $n \geq 4$, in which lines contain at least three points, any maximal subspace is isomorphic to $\mathrm{PG}(n-1,F)$ for some skew field F. Now an easy connectedness argument shows that the same field F coordinatises every maximal subspace.

Theorem 7.11 *Let G be an abstract polar space of rank $n \geq 4$, in which each next-to-maximal subspace is contained in exactly two maximal subspaces. Assume that some maximal subspace is isomorphic to $\mathrm{PG}(n-1,F)$. Then F is commutative, and G is isomorphic to the hyperbolic quadric of rank n over F.*

Proof It is enough to show that F is commutative and that n and F uniquely determine the geometry, since the hyperbolic quadric clearly has the required property.

Rather than prove F commutative, I will show merely that it is isomorphic to its opposite. It suffices to show this when $n = 4$. Take two maximal subspaces meeting in a plane Π, and a point $p \in \Pi$. By the FTPG, both maximal subspaces are isomorphic to $\mathrm{PG}(3,F)$. Now consider the residue of p. This is a projective space, in which there is a plane isomorphic to $\mathrm{PG}(2,F^\circ)$, and a point residue isomorphic to $\mathrm{PG}(2,F)$. Hence $F \cong F^\circ$. The stronger statement that F is commutative is shown by Tits. He observes that the quotient of p has a polarity

interchanging a point and a plane incident with it, and fixing every line incident with both; and this can only happen in a projective 3-space over a commutative field.

Let U_1 and U_2 be disjoint maximal subspaces. Note that they have the same type if n is even, opposite types if n is odd. Let p be any point in neither subspace. Then for $i = 1, 2$, there is a unique maximal subspace W_i containing p and meeting U_i in a hyperplane. Then W_i has the opposite type to U_i, so W_1 and W_2 have the same type if n is even, opposite types if n is odd. Thus, their intersection has codimension congruent to n mod 2. Since $p \in W_1 \cap W_2$, the intersection is at least a line. But their distance in the dual polar space is at least $n-2$, since U_1 and U_2 have distance n; so $W_1 \cap W_2$ is a line L. Clearly L meets both U_1 and U_2.

Each point of U_1 is adjacent to a hyperplane of U_2, and *vice versa*; so U_1 and U_2 are naturally duals. Now the lines joining points of U_1 and U_2 are easily described, and it is not hard to show that the whole geometry is determined. ■

7.5 Recognising quadrics over GF(2)

In this section, we determine the abstract polar spaces with three points on every line. Since we are given information only about points and lines, the BS approach is the natural one. The result here was first found by Shult (assuming a constant number of lines per point) and Seidel (in general), and was a crucial precursor of the Buekenhout–Shult Theorem (Theorem 7.7). Shult and Seidel proved the theorem by induction on the rank: a rank 2 polar space is a generalised quadrangle, and the classification in this case is Theorem 7.3. The elegant direct argument given here is due to Jonathan Hall.

Let G be an abstract polar space with three points per line. We have already seen that the facts that two points lie on at most one line, and that maximal singular subspaces are projective spaces, are proved more easily under this hypothesis than in general. But here is a direct proof of the first assertion. Suppose that the points a and b lie on two lines $\{a,b,x\}$ and $\{a,b,y\}$. Then y is collinear with a and b, and so also with x; so there is a line $\{x,y,z\}$ for some z, and both a and b are joined to z. Any further point is joined to both or neither x and y, and so is joined to z, contradicting (BS2).

Define a graph Γ whose vertices are the points, two vertices being adjacent if they are collinear. The graph has the following property:

(T) every edge $\{x,y\}$ lies in a triangle $\{x,y,z\}$ with the property that any further point is joined to one or all of $\{x,y,z\}$.

This is called the *triangle property*. Shult and Seidel phrased their result as the determination of finite graphs with the triangle property. (The argument just given shows that, in a graph with the triangle property in which no vertex is adjacent to all others, there is a unique triangle with the property specified by (T) containing any edge. Thus, the graph and the polar space determine each other.) The proof given below is not the original argument of Shult and Seidel, which used induction, but is a direct argument due to Jonathan Hall (having the added feature that it works equally well for infinite-dimensional spaces).

Theorem 7.12 *An abstract polar space in which each line contains three points is a quadric over* GF(2).

Proof As noted above, we may assume instead that we have a graph Γ with the triangle property (T), having at least one edge, and having no vertex adjacent to all others. Let X be the vertex set of the graph Γ, and let $F = \mathrm{GF}(2)$. We begin with the vector space $\hat{V}$ of all functions from X to F which are zero everywhere except on a finite set, with pointwise operations. (If X is finite, then V is just the space F^X of all functions from X to F.) Let $\hat{x} \in \hat{V}$ be the characteristic function of the singleton set $\{x\}$. The functions $\hat{x}$, for $x \in X$, form a basis for $\hat{V}$. We define a bilinear form $\hat{b}$ on $\hat{V}$ by setting

$$\hat{b}(\hat{x},\hat{y}) = \begin{cases} 0 & \text{if } x = y \text{ or } x \text{ is joined to } y, \\ 1 & \text{otherwise,} \end{cases}$$

and extending linearly, and a quadratic form $\hat{f}$ by setting $\hat{f}(\hat{x}) = 0$ for all $x \in X$ and extending to $\hat{V}$ by the rule

$$\hat{f}(\mathbf{v}+\mathbf{w}) = \hat{f}(\mathbf{v}) + \hat{f}(\mathbf{w}) + \hat{b}(\mathbf{v},\mathbf{w}).$$

Note that both $\hat{b}$ and $\hat{f}$ are well-defined.

Let R be the *radical* of $\hat{f}$; that is, R is the *subspace*

$$\{\mathbf{v} \in \hat{V} : \hat{f}(\mathbf{v}) = 0,\ \hat{b}(\mathbf{v},\mathbf{w}) = 0 \text{ for all } \mathbf{w} \in \hat{V}\},$$

and set $V = \hat{V}/R$. Then $\hat{b}$ and $\hat{f}$ induce bilinear and quadratic forms b, f on V: for example, we have $f(\mathbf{v}+R) = \hat{f}(\mathbf{v})$ (and this is well-defined, that is, independent of the choice of coset representative). Now let $\overline{x} = \hat{x} + R \in V$.

We claim that the embedding $x \mapsto \overline{x}$ has the required properties; in other words, it is one-to-one; its image is the quadric defined by f; and two vertices are adjacent if and only if the corresponding points of the quadric are orthogonal. We proceed in a series of steps.

Step 1 *Let $\{x,y,z\}$ be a special triangle, as in the statement of the triangle property (T). Then $\overline{x}+\overline{y}+\overline{z}=0$.*

It is required to show that $r=\hat{x}+\hat{y}+\hat{z}\in R$. We have

$$\hat{b}(r,\hat{v})=\hat{b}(\hat{x},\hat{v})+\hat{b}(\hat{y},\hat{v})+\hat{b}(\hat{z},\hat{v})=0$$

for all $v\in X$, by the triangle property; and

$$\hat{f}(r)=\hat{f}(\hat{x})+\hat{f}(\hat{y})+\hat{f}(\hat{z})+\hat{b}(\hat{x},\hat{y})+\hat{b}(\hat{y},\hat{z})+\hat{b}(\hat{z},\hat{x})=0$$

by definition.

Step 2 *The map $x\mapsto\overline{x}$ is one-to-one on X.*

Suppose that $\overline{x}=\overline{y}$. Then $r=\hat{x}+\hat{y}\in R$. Hence $\hat{b}(\hat{x},\hat{y})=0$, and so x is joined to y. Let z be the third vertex of the special triangle containing x and y. Then $\hat{z}=\hat{x}+\hat{y}\in R$ by Step 1, and so z is joined to all other points of X, contrary to assumption.

Step 3 *Any quadrangle is contained in a 3×3 grid.*

Let $\{x,y,z,w\}$ be a quadrangle. Letting $\overline{x+y}=\overline{x}+\overline{y}$, etc., we see that $x+y$ is not joined to z or w, and hence is joined to $z+w$. Similarly, $y+z$ is joined to $w+x$; and the third point in the special triangle through each of these pairs is $x+y+z+w$, completing the grid. (See Fig. 7.1.)

Step 4 *For any $v\in V$, write $v=\sum_{i\in I}\overline{x_i}$, where $x_i\in X$, and the number $m=|I|$ of summands is minimal (for the given v). Then*

(a) $m\leq 3$;

(b) the points x_i are pairwise non-adjacent.

This is the crucial step, and needs four sub-stages.

Substep 4.1 *Assertion (b) is true.*

If $x_i\sim x_j$, we could replace x_i+x_j by the third point x_k of the special triangle, and obtain a shorter expression.

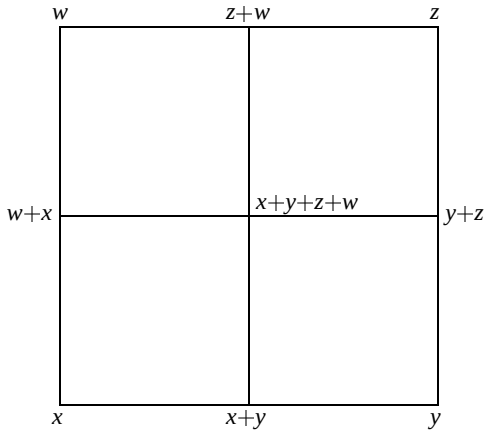


Figure 7.1: A grid

Substep 4.2 *If L is a line on x_1, and y a point of L which is adjacent to x_2, then $y \sim x_i$ for all $i \in I$.*

If not, let $L = \{x_1, y, z\}$, and suppose that $x_i \sim z$. Then x_i is joined to the third point w of the line x_2y. Let u be the third point on x_iw. Then $\bar{z} + \bar{p}_1 + \bar{u} + \bar{x}_i = \bar{x}_2$, and we can replace $\bar{x}_1 + \bar{x}_2 + \bar{x}_i$ by the shorter expression $\bar{z} + \bar{u}$.

Substep 4.3 *There are two points y, z joined to all x_i.*

Each line through x_1 contains a point with this property, by Substep 4.2. It is easily seen that if x_1 lies on a unique line, then one of the points on this line is adjacent to all others, contrary to assumption.

Substep 4.4 $m \leq 3$.

Suppose not. Considering the quadrangles $\{x_1, y, x_2, z\}$ and $\{x_3, y, x_4, z\}$, we find (by Step 3) points a and b with

$$\bar{x}_1 + \bar{y} + \bar{x}_2 + \bar{z} = \bar{a}, \qquad \bar{x}_3 + \bar{y} + \bar{x}_4 + \bar{z} = \bar{b}.$$

But then $\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4 = \bar{a} + \bar{b}$, a shorter expression.

Step 5 *If $v \in V$, $v \neq 0$, and $f(v) = 0$, then $v = \bar{x}$ for some $x \in X$.*

If not then, by Step 4, either $v = \bar{x}+\bar{y}$, or $v = \bar{x}+\bar{y}+\bar{z}$, where points x, y (and z) are (pairwise) non-adjacent. In the second case,

$$f(v) = f(\bar{x}) + f(\bar{y}) + f(\bar{z}) + b(\bar{x},\bar{y}) + b(\bar{y},\bar{z}) + b(\bar{z},\bar{x}) = 0+0+0+1+1+1 = 1.$$

The other case is similar but easier.

Step 6 *$x \sim y$ if and only if $b(\bar{x},\bar{y}) = 0$.*

This is true by definition. ■

7.6 The general case

A weak form of the general classification of polar spaces, by Veldkamp and Tits, can be stated as follows.

Theorem 7.13 *A polar space of type T having finite rank $n \geq 4$ is either classical, or defined by a pseudoquadratic form on a vector space over a division ring of characteristic* 2. ■

I will not attempt to outline the proof of this theorem, but merely make some remarks, including a "definition" of a pseudoquadratic form.

Let V be a vector space over a skew field F of characteristic 2, and σ an anti-automorphism of F satisfying $\sigma^2 = 1$. Let K_0 be the additive subgroup $\{x + x^\sigma\}$ of F, and $K^* = K/K_0$. A function $f : V \to K^*$ is called a *pseudo-quadratic form* relative to σ if there is a σ-sesquilinear form g such that $f(\mathbf{v}) = g(\mathbf{v},\mathbf{v}) \bmod K_0$. Equivalently, f polarises to a σ-Hermitian form f satisfying $(\forall \mathbf{v} \in V)(\exists c \in F)(f(\mathbf{v},\mathbf{v}) = c + c^\sigma)$, that is, a trace-valued form. The function f defines a polar space, consisting of the subspaces of V on which f vanishes (mod K_0). If K_0 is equal to the fixed field of σ, then the same polar space is defined by the Hermitian form g; so we may assume that this is not the case in the second conclusion of Theorem 7.13. For further discussion, see Tits [S].

Tits' result is actually better than indicated: all polar spaces of rank $n \geq 3$ are classified. There are two types of polar spaces of rank 3 which are not covered by Theorem 7.13. The first exists over any non-commutative field, and will be described in the first section of Chapter 8. The other is remarkable in consisting of the only polar spaces whose planes are non-Desarguesian. This type is constructed by Tits from the algebraic groups of type E_6, and again I refer to Tits for the construction, which requires detailed knowledge of these algebraic groups. The

planes actually satisfy a weakening of Desargues' theorem known as the *Moufang condition*, and can be "coordinatised" by certain *alternative division rings* which generalise the *Cayley numbers* or *octonions*.

Of course, the determination of polar spaces of rank 2 (GQs) is a hopeless task! Nevertheless, it is possible to formulate the Moufang condition for generalised quadrangles; and all GQs satisfying the Moufang condition have been determined (by Fong and Seitz in the finite case, Tits and Weiss in general.) This effectively completes the analogy with coordinatisation theorems for projective spaces.

The other geometric achievement of Tits in the 1974 lecture notes is the analogue of the Fundamental Theorem of Projective Geometry:

Theorem 7.14 *Any isomorphism between classical polar spaces of rank at least* 2*, which are not of symplectic or orthogonal type in characteristic* 2*, is induced by a semilinear transformation of the underlying vector spaces.* ■

The reason for the exception will be seen in Section 8.4. As in Section 1.3, this result shows that the automorphism groups of classical polar spaces consist of semilinear transformations modulo scalars. These groups, with some exceptions of small rank, have "large" simple subgroups, just as happened for the automorphism groups of projective spaces in Section 4.6. These groups are the *classical groups*, and are named after their polar spaces: *symplectic*, *orthogonal* and *unitary* groups. For details, see the classic accounts: Dickson [K], Dieudonné [L], and Artin [B], or for more recent accounts Taylor [R], Cameron [10]. In the symplectic or unitary case, the classical group consists of all the linear transformations of determinant 1 preserving the form defining the geometry, modulo scalars. In the orthogonal case, it is sometimes necessary to pass to a subgroup of index 2. (For example, if the polar space is a hyperbolic quadric in characteristic 2, take the subgroup fixing the two families of maximal t.s. subspaces.)

8

The Klein quadric and triality

Low-dimensional hyperbolic quadrics possess a remarkably rich structure; the Klein quadric in 5-space encodes a projective 3-space, and the triality quadric in 7-space possesses an unexpected threefold symmetry. The contents of this chapter can be predicted from the diagrams of these geometries, since D_3 is isomorphic to A_3, and D_4 has an automorphism of order 3.

8.1 The Pfaffian

The determinant of a skew-symmetric matrix is a square. This can be seen in small cases by direct calculation:

$$\det\begin{pmatrix} 0 & a_{12} \\ -a_{12} & 0 \end{pmatrix} = a_{12}^2,$$

$$\det\begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{pmatrix} = (a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23})^2.$$

Theorem 8.1 *(a) The determinant of a skew-symmetric matrix of odd size is zero.*

(b) There is a unique polynomial $\mathrm{Pf}(A)$ *in the indeterminates* a_{ij} *for* $1 \le i < j \le 2n$*, having the properties*

(i) if A *is a skew-symmetric* $2n \times 2n$ *matrix with* (i, j) *entry* a_{ij} *for* $1 \le i < j \le 2n$*, then*

$$\det(A) = \mathrm{Pf}(A)^2;$$

(ii) $\mathrm{Pf}(A)$ *contains the term* $a_{12}a_{34}\cdots a_{2n-1\,2n}$ *with coefficient* $+1$.

Proof We begin by observing that, if A is a skew-symmetric matrix, then the form B defined by

$$B(x,y) = xAy^\top$$

is an alternating bilinear form. Moreover, B is non-degenerate if and only if A is non-singular: for $xAy^\top = 0$ for all y if and only if $xA = 0$. We know that there is no non-degenerate alternating bilinear form on a space of odd dimension; so (a) is proved.

We know also that, if A is singular, then $\det(A) = 0$, whereas if A is non-singular, then there exists an invertible matrix P such that

$$PAP^\top = \mathrm{diag}\left(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\right),$$

so that $\det(A) = \det(P)^{-2}$. Thus, $\det(A)$ is a square in either case.

Now regard a_{ij} as being indeterminates over the field F; that is, let $K = F(a_{ij} : 1 \le i < j \le 2n)$ be the field of fractions of the polynomial ring in $n(2n-1)$ variables over F. If A is the skew-symmetric matrix with entries a_{ij} for $1 \le i < j \le 2n$, then as we have seen, $\det(A)$ is a square in K. It is actually the square of a polynomial. (For the polynomial ring is a unique factorisation domain; if $\det(A) = (f/g)^2$, where f and g are polynomials with no common factor, then $\det(A)g^2 = f^2$, and so f^2 divides $\det(A)$; this implies that g is a unit.) Now $\det(A)$ contains a term

$$a_{12}^2 a_{34}^2 \cdots a_{2n-1\,2n}^2$$

corresponding to the permutation

$$(12)(34)\cdots(2n-1\,2n),$$

and so by choice of sign in the square root we may assume that (ii)(b) holds. Clearly the polynomial $\mathrm{Pf}(A)$ is uniquely determined.

The result for arbitrary skew-symmetric matrices is now obtained by specialisation (that is, substituting values from F for the indeterminates a_{ij}). ■

Exercises

1. A *one-factor* on the set $\{1,2,\ldots,2n\}$ is a partition F of this set into n subsets of size 2. We represent each 2-set$\{i,j\}$ by the ordered pair (i,j) with $i < j$. The *crossing number* $\chi(F)$ of the one-factor F is the number of pairs $\{(i,j),(k,l)\}$ of sets in F for which $i < k < j < l$.

(a) Let $\mathcal{F}_n$ be the set of one-factors on the set $\{1,2,\ldots,2n\}$. What is $|\mathcal{F}_n|$?

(b) Let $A=(a_{ij})$ be a skew-symmetric matrix of order $2n$. Prove that

$$\mathrm{Pf}(A)=\sum_{F\in\mathcal{F}_n}(-1)^{\chi(F)}\prod_{(i,j)\in F}a_{ij}.$$

2. Show that, if A is a skew-symmetric matrix and P any invertible matrix, then

$$\mathrm{Pf}(PAP^\top)=\det(P)\cdot\mathrm{Pf}(A).$$

Hint: We have $\det(PAP^\top)=\det(P)^2\det(A)$, and taking the square root shows that $\mathrm{Pf}(PAP^\top)=\pm\det(P)\,\mathrm{Pf}(A)$; it is enough to justify the positive sign. Show that it suffices to consider the 'standard' skew-symmetric matrix

$$A=\mathrm{diag}\left(\begin{pmatrix}0&1\\-1&0\end{pmatrix},\ldots,\begin{pmatrix}0&1\\-1&0\end{pmatrix}\right),$$

In this case, show that the $(2n-1,2n)$ entry in $PAP^\top$ contains the term $p_{2n-1\,2n-1}p_{2n\,2n}$, so that $\mathrm{Pf}(PAP^\top)$ contains the diagonal entry of $\det(P)$ with sign $+1$.

3. Show that any linear transformation of a vector space fixing a symplectic form (a non-degenerate alternating bilinear form) has determinant 1.

8.2 The Klein correspondence

We begin by describing an abstract polar space which appears not to be of classical type. Let F be a skew field, and consider the geometry $\mathcal{G}$ defined from $\mathrm{PG}(3,F)$ as follows:

- the POINTs of $\mathcal{G}$ are the lines of $\mathrm{PG}(3,F)$;
- the LINEs of $\mathcal{G}$ are the plane pencils (incident point-plane pairs);
- the PLANEs of $\mathcal{G}$ are of two types: the points, and the planes.

A POINT and LINE are incident if the line belongs to the plane pencil (i.e., is incident with both the point and the plane). A LINE and PLANE are incident if the point or plane is one of the elements of the incident pair; and incidence between a POINT and a PLANE is the usual incidence in $\mathrm{PG}(3,F)$.

If a PLANE is a plane Π, then the POINTs and LINEs of this PLANE correspond to the lines and points of Π; so the residue of the plane is isomorphic to the

dual of Π, namely, $\mathrm{PG}(2,F^\circ)$. On the other hand, if a PLANE is a point p, then the POINTs and LINEs of this PLANE are the lines and planes through p, so its residue is the residue of p in $\mathrm{PG}(3,F)$, namely $\mathrm{PG}(2,F)$. Thus (PS1) holds. (Note that, if F is not isomorphic to its opposite, then the space contains non-isomorphic planes, something which cannot happen in a classical polar space.)

Axiom (PS2) is clear. Consider (PS3). Suppose that the PLANE in question is a plane Π, and the POINT not incident with it is a line L. Then $L\cap\Pi$ is a point p; the set of POINTs of Π collinear with L is the plane pencil defined by p and Π (which is a LINE), and the union of the LINEs joining them to L consists of all lines through p (which is a PLANE), as required. The other case is dual.

Finally, if the point p and plane Π are non-incident, then the PLANEs they define are disjoint, proving (PS4).

Note that any LINE is incident with just two PLANEs, one of each type; so, if the polar space is classical, it must be a hyperbolic quadric in $\mathrm{PG}(5,F)$. We now show that, if F is commutative, it is indeed this quadric in disguise! (For non-commutative fields, this is one of the exceptional rank 3 polar spaces mentioned in Section 7.6.)

The skew-symmetric matrices of order 4 over F form a vector space of rank 6, with $x_{12},\ldots,x_{34}$ as coordinates. The Pfaffian is a quadratic form on this vector space, which vanishes precisely on the singular matrices. So, projectively, the singular matrices form a quadric Q in $\mathrm{PG}(5,F)$, the so-called *Klein quadric*. From the form of the Pfaffian, we see that this quadric is hyperbolic — but in fact this will become clear geometrically.

Any skew-symmetric matrix has even rank. In our case, a non-zero singular skew-symmetric matrix A has rank 2, and so can be written in the form

$$A = X(\mathbf{v},\mathbf{w}) := \mathbf{v}^\top\mathbf{w} - \mathbf{w}^\top\mathbf{v}$$

for some vectors $\mathbf{v},\mathbf{w}$. Replacing these two vectors by linear combinations $\alpha\mathbf{v}+\beta\mathbf{w}$ and $\gamma\mathbf{v}+\delta\mathbf{w}$ multiplies A by a factor $\alpha\delta-\beta\gamma$ (which is just the determinant of the transformation). So we have a map from the line of $\mathrm{PG}(3,F)$ spanned by $\mathbf{v}$ and $\mathbf{w}$ to the point of the Klein quadric spanned by $X(\mathbf{v},\mathbf{w})$. This map is a bijection: we have seen that it is onto, and the matrix determines the line as its row space.

This bijection has the properties predicted by our abstract treatment. Most important,

> two points of the Klein quadric are perpendicular if and only if the corresponding lines intersect.

To prove this, note that two points are perpendicular if and only if the line joining them lies in Q. Now, if two lines intersect, we can take them to be $\langle \mathbf{u}, \mathbf{v} \rangle$ and $\langle \mathbf{u}, \mathbf{w} \rangle$; and we have

$$\alpha(\mathbf{u}^\top \mathbf{v} - \mathbf{v}^\top \mathbf{u}) + \beta(\mathbf{u}^\top \mathbf{w} - \mathbf{w}^\top \mathbf{u}) = \mathbf{u}^\top(\alpha\mathbf{v} + \beta\mathbf{w}) - (\alpha\mathbf{v} + \beta\mathbf{w})^\top \mathbf{u},$$

so the line joining the corresponding points lies in the quadric. Conversely, if two lines are skew, then they are $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ and $\langle \mathbf{v}_3, \mathbf{v}_4 \rangle$, where $\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$ is a basis; then the matrix

$$\mathbf{v}_1^\top \mathbf{v}_2 - \mathbf{v}_2^\top \mathbf{v}_1 + \mathbf{v}_3^\top \mathbf{v}_4 - \mathbf{v}_4^\top \mathbf{v}_3$$

has rank 4, and is a point on the line not on Q.

Hence the planes on the quadric correspond to maximal families of pairwise intersecting lines, of which there are two types: all lines through a fixed point; and all lines in a fixed plane. Moreover, the argument in the preceding paragraph shows that lines on Q do indeed correspond to plane pencils of lines in $\mathrm{PG}(3, F)$. This completes the identification.

Exercise

1. This exercise gives the promised identification of $\mathrm{PSL}(4,2)$ with the alternating group A_8.

Let V be the vector space of rank 6 over $\mathrm{GF}(2)$ consisting of the binary words of length 8 having even weight modulo the subspace Z consisting of the all-zero and all-1 words. Show that the function

$$f(\mathbf{v} + Z) = \tfrac{1}{2}\mathrm{wt}(\mathbf{v}) \pmod 2$$

is well-defined and is a quadratic form of rank 3 on V, whose zeros form the Klein quadric Q. Show that the symmetric group S_8 interchanges the two families of planes on Q, the subgroup fixing the two families being the alternating group A_8.

Use the Klein correspondence to show that A_8 is embedded as a subgroup of $\mathrm{PGL}(4,2) = \mathrm{PSL}(4,2)$. By calculating the orders of these groups, show that equality holds.

Remark The isomorphism between $\mathrm{PSL}(4,2)$ and A_8 can be used to give a solution to Kirkman's *schoolgirl problem*. This problem asks for a schedule for fifteen schoolgirls to walk in five groups of three every day for seven days, subject to the

requirement that any two girls walk together in a group exactly once during the week.

The 7×5 groups of girls are thus the blocks of a 2-$(15,3,1)$ design. We will take this design to consist of the points and lines of $\mathrm{PG}(3,2)$. The problem is then to find a 'parallelism' or 'resolution', a partition of the lines into seven 'parallel classes' each consisting of five pairwise disjoint lines.

One way to find a parallel class is to consider the underlying vector space $V(4,2)$ as a vector space of rank 2 over $\mathrm{GF}(4)$. The five 'points' or rank 1 subspaces over $\mathrm{GF}(4)$ become five pairwise disjoint lines when we restrict the scalars to $\mathrm{GF}(2)$. Scalar multiplication by a primitive element of $\mathrm{GF}(4)$ is an automorphism of order 3, fixing all five lines, and commuting with a subgroup $\mathrm{SL}(2,4) \cong A_5$. Moreover, if two such automorphisms oof order 3 have a common fixed line, then they generate a $\{2,3\}$-group, since the stabiliser of a line in $\mathrm{GL}(4,2)$ is a $\{2,3\}$-group.

Now, in A_8, an element of order 3 commuting with a subgroup isomorphic to A_5 is necessarily a 3-cycle. Two 3-cycles generate a $\{2,3\}$-group if and only if their supports intersect in 0 or 2 points. So we require a set of seven 3-subsets of $\{1,\ldots,8\}$, any two of which meet in one point. The lines of $\mathrm{PG}(2,2)$ (omitting one point) have this property.

8.3 Some dualities

We have interpreted points of the Klein quadric in $\mathrm{PG}(3,F)$. What about the points off the quadric?

Theorem 8.2 *There is a bijection from the set of points p outside the Klein quadric Q to symplectic structures on $\mathrm{PG}(3,F)$, with the property that a point of Q perpendicular to p translates under the Klein correspondence to a totally isotropic line for the symplectic geometry.*

Proof A point $p \notin Q$ is represented by a skew-symmetric matrix A which has non-zero Pfaffian (and hence is invertible), up to a scalar multiple. The matrix defines a symplectic form b, by the rule

$$b(\mathbf{v},\mathbf{w}) = \mathbf{v}A\mathbf{w}^\top.$$

We must show that a line is t.i. with respect to this form if and only if the corresponding point of Q is perpendicular to p.

Let A be a non-singular skew-symmetric 4×4 matrix over a field F. By direct calculation, we show that the following assertions are equivalent, for any vectors $\mathbf{v},\mathbf{w}\in F^4$:

(a) $X(\mathbf{v},\mathbf{w})=\mathbf{v}^\top\mathbf{w}-\mathbf{w}^\top\mathbf{v}$ is orthogonal to A, with respect to the bilinear form obtained by polarising the quadratic form $Q(X)=\mathrm{Pf}(X)$;

(b) $\mathbf{v}$ and $\mathbf{w}$ are orthogonal with respect to the symplectic form with matrix $A^\dagger$, that is, $\mathbf{v}A^\dagger\mathbf{w}^\top=0$.

Here the matrices A and $A^\dagger$ are given by

$$A=\begin{pmatrix} 0 & a_{12} & a_{13} & a_{14}\\ -a_{12} & 0 & a_{23} & a_{24}\\ -a_{13} & -a_{23} & 0 & a_{34}\\ -a_{14} & -a_{24} & -a_{34} & 0\end{pmatrix},\qquad A^\dagger=\begin{pmatrix} 0 & a_{34} & -a_{24} & a_{23}\\ -a_{34} & 0 & a_{14} & -a_{13}\\ a_{24} & -a_{14} & 0 & a_{12}\\ -a_{23} & a_{13} & -a_{12} & 0\end{pmatrix}.$$

Note that, if A is the matrix of the standard symplectic form, then so is $A^\dagger$. In general, the map taking the point outside the quadric spanned by A to the symplectic form with matrix $A^\dagger$ is the one asserted in the theorem. ∎

Let $\mathcal{G}_1$ be the symplectic GQ over F, and $\mathcal{G}_2$ the orthogonal GQ associated with the quadric $\mathbf{v}^\perp\cap\mathcal{Q}$, where $\mathcal{Q}$ is the Klein quadric and $\langle\mathbf{v}\rangle\notin\mathcal{Q}$. (Note that any non-singular quadratic form of rank 2 in 5 variables is equivalent to $\alpha x_0^2+x_1x_2+x_3x_4$ for some $\alpha\neq 0$; so any two such forms are equivalent up to scalar multiple, and define the same GQ.) We have defined a map from points of $\mathcal{G}_2$ to lines of $\mathcal{G}_1$. Given any point p of $\mathcal{G}_1$, the lines of $\mathcal{G}_1$ containing p form a plane pencil in $\mathrm{PG}(3,F)$, and so translate into a line of $\mathcal{G}_2$. Thus we have shown:

Theorem 8.3 *For any field F, the symplectic GQ in $\mathrm{PG}(3,F)$ and the orthogonal GQ in $\mathrm{PG}(4,F)$ are dual.* ∎

Now let F be a field which has a Galois extension K of degree 2 and σ the Galois automorphism of K over F. With the extension K/F we can associate two GQs:

$\mathcal{G}_1'$: the unitary GQ in $\mathrm{PG}(3,K)$, defined by the Hermitian form

$$x_1y_2^\sigma+x_2y_1^\sigma+x_3y_4^\sigma+x_4y_3^\sigma;$$

$\mathcal{G}_2'$: the orthogonal GQ in $\mathrm{PG}(5,F)$ defined by the quadratic form

$$x_1x_2+x_3x_4+\alpha x_5^2+\beta x_5x_6+\gamma x_6^2,$$

where $\alpha x^2+\beta x+\gamma$ is an irreducible quadratic over F which splits in K.

Theorem 8.4 *The two GQs $\mathcal{G}_1'$ and $\mathcal{G}_2'$ defined above are dual.*

Proof This is proved by "twisting the Klein correspondence". In outline, we take the Klein correspondence over K, and change coordinates on the quadric so that restriction of scalars to F gives the geometry $\mathcal{G}_2'$, rather than the Klein quadric over F; then show that the corresponding set of lines in $\mathrm{PG}(3,K)$ are those which are totally isotropic with respect to a Hermitian form. ■

Exercises

1. Prove the assertion about A and $A^\dagger$ in the proof of Theorem 8.2.

Let $\mathcal{Q}$ be a hyperbolic quadric of rank n. If v is a non-singular vector, then the quadric $v^\perp \cap \mathcal{Q} = \mathcal{S}$ has the property

- $\mathcal{S}$ meets every maximal subspace E of $\mathcal{Q}$ in a hyperplane of E.

We call a set $\mathcal{S}$ satisfying this condition *special*. The point of the next three exercises is to investigate whether special sets are necessarily quadrics of the form $v^\perp \cap \mathcal{Q}$.

2. Consider the case $n = 2$. Let the rank 4 vector space be the space of all 2×2 matrices over F, and let the quadratic form be the determinant.

(a) Show that the map

$$\langle X \rangle \mapsto (\mathrm{Ker}(X), \mathrm{Im}(X))$$

induces a bijection between the point set of the quadric $\mathcal{Q}$ and $P \times P$, where P is the projective line over F.

(b) If A is a non-singular matrix, show that

$$A^\perp = \{\langle X \rangle \in \mathcal{Q} : \mathrm{Ker}(X) \cdot A = \mathrm{Im}(X)\},$$

which corresponds under this bijection to the set $\{(p, p \cdot A) : p \in P\}$.

(c) Show that, if π is any permutation of P, then $\{(p, \pi(p)) : p \in P\}$ is a special set; and all special sets have this form.

(d) Deduce that every special set is a quadric if and only if $|F| \leq 3$.

3. Consider the case $n = 3$. Take $\mathcal{Q}$ to be the Klein quadric. Show that the Klein correspondence maps the special set $\mathcal{S}$ to a set S of lines of $\mathrm{PG}(3,F)$ with the property that the set of lines of S through any point of p, or the set of lines of S in any plane Π, is a plane pencil. Show that the correspondence $p \mapsto \Pi$ of $\mathrm{PG}(3,F)$, where the set of lines of S containing p and the set contained in Π are equal, is a symlectic polarity with S as its set of absolute lines. Deduce that S is the set of lines of a symplectic GQ in $\mathrm{PG}(3,F)$, and hence that $\mathcal{S}$ is a quadric.

4. Prove by induction on n that, for $n \geq 3$, any special set is a quadric. (See Cameron and Kantor [12] for a crib.)

8.4 Dualities of symplectic quadrangles

A field of characteristic 2 is said to be *perfect* if every element is a square. A finite field of characteristic 2 is perfect, since the multiplicative group has odd order.

If F has characteristic 2, then the map $x \mapsto x^2$ is a homomorphism, since

$$\begin{aligned}(x+y)^2 &= x^2+y^2,\\ (xy)^2 &= x^2y^2,\end{aligned}$$

and is one-to-one. Hence F is perfect if and only if this map is an automorphism.

Theorem 8.5 *Let F be a perfect field of characteristic 2. Then there is an isomorphism between the symplectic polar space of rank n over F, and the orthogonal polar space of rank n defined by a quadratic form in $2n+1$ variables.*

Proof Let V be a vector space of rank $2n+1$ carrying a non-singular quadratic form f of rank n. By polarising f, we get an alternating bilinear form b, which cannot be non-degenerate; its radical $R = V^\perp$ is of rank 1, and the restriction of f to it is the germ of f.

Let W_0 be a totally singular subspace of V. Then $W = W_0 + R$ is a totally isotropic subspace of the non-degenerate symplectic space V/R. So we have an incidence-preserving injection $\theta : W_0 \mapsto (W_0+R)/R$ from the orthogonal polar space to the symplectic. We have to show that θ is onto.

So let W/R be t.i. This means that W itself is t.i. for the form b; but $R \subseteq W$, so W is not t.s. for f. However, on W, we have

$$\begin{aligned}f(\mathbf{w}_1+\mathbf{w}_2) &= f(\mathbf{w}_1)+f(\mathbf{w}_2),\\ f(\alpha\mathbf{w}) &= \alpha^2 f(\mathbf{w}),\end{aligned}$$

so f is semilinear on W. Thus, the kernel of f is a hyperplane W_0 of W. The space W_0 is t.s., and $W_0 + R = W$; so W_0 maps onto W/R under θ. ∎

Now consider the case $n = 2$. We have an isomorphism between the symplectic and orthogonal quadrangles, by Theorem 8.5, and a duality, by Theorem 8.3. So:

Theorem 8.6 *The symplectic generalised quadrangle over a perfect field of characteristic* 2 *is self-dual.* ∎

When is there a polarity?

Theorem 8.7 *Let F be a perfect field of characteristic* 2*. Then the symplectic GQ over F has a polarity if and only if F has an automorphism σ satisfying*

$$\sigma^2 = 2,$$

where 2 *denotes the automorphism $x \mapsto x^2$.*

Proof For this, we cannot avoid using coordinates! We take the vector space F^4 with the standard symplectic form

$$b((x_1,x_2,x_3,x_4),(y_1,y_2,y_3,y_4)) = x_1y_2 + x_2y_1 + x_3y_4 + x_4y_3.$$

(Remember that the characteristic is 2.) The Klein correspondence takes the line spanned by (x_1,x_2,x_3,x_4) and (y_1,y_2,y_3,y_4) to the point with coordinates z_{ij}, $1 \leq i < j \leq 4$, where $z_{ij} = x_iy_j + x_jy_i$; this point lies on the quadric with equation

$$z_{12}z_{34} + z_{13}z_{24} + z_{14}z_{23} = 0,$$

and (if the line is t.i.) also on the hyperplane $z_{12} + z_{34} = 0$. If we factor out the subspace spanned by the point with $z_{12} = z_{34} = 1$, $z_{ij} = 0$ otherwise, and use coordinates $(z_{13},z_{24},z_{14},z_{23})$, we obtain a point of the symplectic space; the map δ from lines to points is the duality previously defined.

To compute the image of a point $p = (a_1,a_2,a_3,a_4)$ under the duality, take two t.i. lines through this point and calculate their images. If a_1 and a_4 are non-zero, we can use the lines joining p to the points $(a_1,a_2,0,0)$ and $(0,a_4,a_1,0)$; the images are $(a_1a_3,a_2a_4,a_1a_4,a_2a_3)$ and $(a_1^2,a_4^2,0,a_1a_2+a_3a_4)$. Now the image of the line joining these points is found to be the point $(a_1^2,a_2^2,a_3^2,a_4^2)$. The same formula is found in all cases. So δ^2 is the collineation induced by the field automorphism $x \mapsto x^2$, or 2 as we have called it.

Suppose that there is a field automorphism σ with $\sigma^2 = 2$, and let $\theta = \sigma^{-1}$; then $(\delta\theta)^2$ is the identity, so $\delta\theta$ is a polarity.

Conversely, suppose that there is a polarity. By Theorem 7.14, any collineation g is induced by the product of a linear transformation and a uniquely defined field automorphism $\theta(g)$. Now any duality has the form δg for some collineation g; and

$$\theta((\delta g)^2) = 2\theta(g)^2.$$

So, if δg is a polarity, then $2\theta(g)^2 = 1$, whence $\sigma = \theta(g)^{-1}$ satisfies $\sigma^2 = 2$. ∎

In the case where F is a finite field $\mathrm{GF}(2^m)$, the automorphism group of F is cyclic of order m, generated by 2; and so there is a solution of $\sigma^2 = 2$ if and only if m is odd. We conclude that the symplectic quadrangle over $\mathrm{GF}(2^m)$ has a polarity if and only if m is odd.

We now examine the set of *absolute* points and lines (i.e., those incident with their image). A *spread* is a set S of lines such that every point lies on a unique line of S. Dually, an *ovoid* in a GQ is a set O of points with the property that any line contains a unique point of O. Note that this is quite different from the definition of an ovoid in $\mathrm{PG}(3,F)$ given in Section 4.4; but there is a connection, as we will see.

Proposition 8.8 *The set of absolute points of a polarity of a GQ is an ovoid, and the set of absolute lines is a spread.*

Proof Let δ be a polarity. No two absolute points are collinear. For, if x and y are absolute points lying on the line L, then x, y and $L\delta$ would form a triangle.

Suppose that the line L contains no absolute point. Then L is not absolute, so $L\delta \notin L$. Thus, there is a unique line M containing $L\delta$ and meeting L. Then $M\delta \in L$, so $M\delta$ is not absolute. But L meets M, so $L\delta$ and $M\delta$ are collinear; hence $L\delta, M\delta$ and $L \cap M$ form a triangle.

The second statement is dual. ∎

Theorem 8.9 *The set of absolute points of a polarity of a symplectic GQ in* $\mathrm{PG}(3,F)$ *is an ovoid in* $\mathrm{PG}(3,F)$.

Proof Let σ be the polarity of the GQ $\mathcal{G}$, and $\perp$ the polarity of the projective space defining the GQ. By the last result, the set O of absolute points of σ is an ovoid in $\mathcal{G}$. This means that the t.i. lines are tangents to O, and the t.i. lines through a point of O form a plane pencil. So we have to prove that any other line of the projective space meets O in 0 or 2 points.

Let X be a hyperbolic line, p a point of $X \cap O$, and $p^\sigma = L$. Then L meets the hyperbolic line $L^\perp$ in a point q. Let $q^\sigma = M$. Since $q \in L$, we have $p \in M$; so M also meets $X^\perp$, in a point r. Let $N = r^\sigma$. Then $q \in N$, so N meets X. Also, N meets O in a point s. The line s^σ contains s and $N^\sigma = r$. So s is on two lines meeting $X^\perp$, whence $s \in X$. So, if $|X \cap O| \geq 1$, then $|X \cap O| \geq 2$.

Now let p' be another point of $X \cap O$, and define L' and q' as before. Let K be the line pq'. Then $p \in K$, so $p^\sigma = L$ contains $x = K^\sigma$. Also, K meets L', so x is collinear with p'. But the only point of L collinear with p' is q. So $x = q$, independent of p'. This means that there is only one point $p' \neq p$ in $X \cap O$, and this set has cardinality 2. ■

Remark Over finite fields, any ovoid in a symplectic GQ is an ovoid in the ambient projective 3-space. This is false for infinite fields. (See Exercises 2 and 3.)

Hence, if F is a perfect field of characteristic 2 in which $\sigma^2 = 2$ for some automorphism σ, then $\mathrm{PG}(3,F)$ possesses symplectic ovoids and spreads. These give rise to inversive planes and to translation planes, as described in Sections 4.1 and 4.4. For finite fields F, these are the only known ovoids other than elliptic quadrics.

Exercises

1. Suppose that the points and lines of a GQ are all the points and some of the lines of $\mathrm{PG}(3,F)$. Prove that the lines through any point form a plane pencil, and deduce that the GQ is symplectic.

2. Prove that an ovoid O in a symplectic GQ over the finite field $\mathrm{GF}(q)$ is an ovoid in $\mathrm{PG}(3,q)$. [Hint: as in Theorem 8.3.5, it suffices to prove that any hyperbolic line meets O in 0 or 2 points. Now, if X is a hyperbolic line with $X \cap O \neq \emptyset$, then $X^\perp \cap O = \emptyset$, so at most half of the $q^2(q^2+1)$ hyperbolic lines meet O. Take any $N = \frac{1}{2}q^2(q^2+1)$ hyperbolic lines including all those meeting O, and let n_i of the chosen lines meet O in i points. Prove that $\sum n_i = N$, $\sum i n_i = 2N$, $\sum i(i-1)n_i = 2N$.]

3. Prove that, for any infinite field F, there is an ovoid of the symplectic quadrangle over F which is not an ovoid of the embedding projective space.

8.5 Reguli and spreads

We met in Section 4.1 the concepts of a regulus in $\mathrm{PG}(3,F)$ (the set of common transversals to three pairwise skew lines), a spread (a set of pairwise skew lines covering all the points), a bispread (a spread containing a line of each plane), and a regular spread (a spread containing the regulus through any three of its lines). We now translate these concepts to the Klein quadric.

Theorem 8.10 *Under the Klein correspondence,*

(a) a regulus corresponds to a conic, the intersection if $\mathcal{Q}$ with a non-singular plane Π, and the opposite regulus to the intersection of $\mathcal{Q}$ with $\Pi^\perp$;

(b) a bispread corresponds to an ovoid, a set of pairwise non-perpendicular points meeting every plane on $\mathcal{Q}$;

(c) a regular spread corresponds to the ovoid $\mathcal{Q}\cap W^\perp$, where W is a line disjoint from $\mathcal{Q}$.

Proof (a) Take three pairwise skew lines. They translate into three pairwise non-perpendicular points of $\mathcal{Q}$, which span a non-singular plane Π (so that $\mathcal{Q}\cap\Pi$ is a conic C). Now $\Pi^\perp$ is also a non-singular plane, and $\mathcal{Q}\cap\Pi^\perp$ is a conic C', consisting of all points perpendicular to the three given points. Translating back, C' corresponds to the set of common transversals to the three given lines. This set is a regulus, and is opposite to the regulus spanned by the given lines (corresponding to C).

(b) This is straightforward translation. Note, incidentally, that a spread (or a cospread) corresponds to what might be called a "semi-ovoid", were it not that this term is used for a different concept: that is, a set of pairwise non-perpendicular points meeting every plane in one family on $\mathcal{Q}$.

(c) A regular spread is "generated" by any four lines not contained in a regulus, in the sense that it is obtained by repeatedly adjoining all the lines in a regulus through three of its lines. On $\mathcal{Q}$, the four given lines translate into four points, and the operation of generation leaves us within the 3-space they span. This 3-space has the form $W^\perp$ for some line W; and no point of $\mathcal{Q}$ can be perpendicular to every point of such a 3-space. ■

Note that a line disjoint from $\mathcal{Q}$ is anisotropic; such lines exist if and only if there is an irreducible quadratic over F, that is, F is not quadratically closed. (We

saw earlier the construction of regular spreads: if K is a quadratic extension of F, take the rank 1 subspaces of a rank 2 vector space over K, and restrict scalars to F.)

Thus a bispread is regular if and only if the corresponding ovoid is contained in a 3-space section of $\mathcal{Q}$. A bispread whose ovoid lies in a 4-space section of $\mathcal{Q}$ is called *symplectic*, since its lines are totally isotropic with respect to some symplectic form (by the results of Section 8.3). An open problem is to find a simple structural test for symplectic bispreads (resembling the characterisation of regular spreads in terms of reguli).

We also saw in Section 4.1 that spreads of lines in projective space give rise to translation planes; and regular spreads give Desarguesian (or Pappian) planes. Another open problem is to characterise the translation planes arising from symplectic spreads or bispreads.

8.6 Triality

Now we increase the rank by 1, and let $\mathcal{Q}$ be a hyperbolic quadric in $\mathrm{PG}(7,F)$, defined by a quadratic form of rank 4. The maximal t.s. subspaces have dimension 3, and are called *solids*; as usual, they fall into two families $\mathcal{M}_1$ and $\mathcal{M}_2$, so that two solids in the same family meet in a line or are disjoint, while two solids in different families meet in a plane or a point. Any t.s. plane lies in a unique solid of each type. Let $\mathcal{P}$ and $\mathcal{L}$ be the sets of points and lines.

Consider the geometry defined as follows.

- The POINTs are the elements of $\mathcal{M}_1$.
- The LINEs are the elements of $\mathcal{L}$.
- The PLANEs are incident pairs (p,M), $p \in \mathcal{P}$, $M \in \mathcal{M}_2$.
- The SOLIDs are the elements of $\mathcal{P} \cup \mathcal{M}_2$.

Incidence is defined as follows. Between POINTs, LINEs and SOLIDs, it is as in the quadric, with the additional rule that the POINT M_1 and SOLID M_2 are incident if they intersect in a plane. The PLANE (p,M) is incident with all those varieties incident with both p and M.

Proposition 8.11 *The geometry just described is an abstract polar space in which any PLANE is incident with just two SOLIDs.*

Proof We consider the axioms in turn.

(P1): Consider, for example, the SOLID $M \in \mathcal{M}_2$. The POINTs incident with M are bijective with the planes of M; the LINEs are the lines of M; the PLANEs are pairs (p,M) with $p \in M$, and so are bijective with the points of M. Incidence is defined so as to make the subspaces contained in M a projective space isomorphic to the dual of M.

For the SOLID $p \in \mathcal{P}$, the argument is a little more delicate. The geometry $p^\perp/p$ is a hyperbolic quadric in $\mathrm{PG}(5,F)$, that is, the Klein quadric; the POINTs, LINEs and PLANEs incident with p are bijective with one family of planes, the lines, and the other family of planes on the quadric; and hence (by the Klein correspondence) with the points, lines and planes of $\mathrm{PG}(3,F)$.

The other cases are easier.

(P2) is trivial, (P3) routine, and (P4) is proved by observing that if $p \in \mathcal{P}$ and $M \in \mathcal{M}_2$ are not incident, then no POINT can be incident with both.

Finally, the SOLIDs containing the PLANE (p,M) are p and M only. ■

So the new geometry we constructed is itself a hyperbolic quadric in $\mathrm{PG}(7,F)$, and hence isomorphic to the original one. This implies the existence of a map τ which carries $\mathcal{L}$ to itself and $\mathcal{P} \to \mathcal{M}_1 \to \mathcal{M}_2 \to \mathcal{P}$. This map is called a *triality* of the quadric, by analogy with dualities of projective spaces.

It is more difficult to describe trialities in coordinates. An algebraic approach must wait until Chapter 10.

Exercise

1. Prove the Buekenhout-Shult property for the geometry constructed in this section. That is, let $M \in \mathcal{M}_1$, $L \in \mathcal{L}$, and suppose that L is not incident with M; prove that either all members of $\mathcal{M}_1$ containing L meet M in a plane, or just one does, depending on whether L is disjoint from M or not.

8.7 An example

In this section we apply triality to the solution of a combinatorial problem first posed and settled by Breach and Street [2]. Our approach follows Cameron and Praeger [13].

Consider the set of planes of $\mathrm{AG}(3,2)$. They form a 3-$(8,4,1)$ design, that is, a collection of fourteen 4-subsets of an 8-set, any three points contained in exactly one of them. There are $\binom{8}{4} = 70$ 4-subsets altogether; can they be partitioned into

five copies of $\mathrm{AG}(3,2)$? The answer is "no", as has been known since the time of Cayley. (In fact, there cannot be more than two disjoint copies of $\mathrm{AG}(3,2)$ on an 8-set; a construction will be given in the next chapter.) Breach and Street asked: what if we take a 9-set? This has $\binom{9}{4} = 126$ 4-subsets, and can conceivably be partitioned into nine copies of $\mathrm{AG}(3,2)$, each omitting one point. They proved:

Theorem 8.12 *There are exactly two non-isomorphic ways to partition the* 4*-subsets of a* 9*-set into nine copies of* $\mathrm{AG}(3,2)$*. Both admit* 2*-transitive groups.*

Proof First we construct the two examples.

1. Regard the 9-set as the projective line over $\mathrm{GF}(8)$. If any point is designated as the point at infinity, the remaining points form an affine line over $\mathrm{GF}(8)$, and hence (by restricting scalars) an affine 3-space over $\mathrm{GF}(2)$. We take the fourteen planes of this affine 3-space as one of our designs, and perform the same construction for each point to obtain the desired partition. This partition is invariant under the group $\mathrm{P\Gamma L}(2,8)$, of order $9 \cdot 8 \cdot 7 \cdot 3 = 1512$. The automorphism group is the stabiliser of the object in the symmetric group; so the number of partitions isomorphic to this one is the index of this group in S_9, which is $9!/1512 = 240$.

2. Alternatively, the nine points carry the structure of affine plane over $\mathrm{GF}(3)$. Identifying one point as the origin, the structure is a rank 2 vector space over $\mathrm{GF}(3)$. Put a symplectic form b on the vector space. Now there are six 4-sets which are symmetric differences of two lines through the origin, and eight 4-sets of the form $\{\mathbf{v}\} \cup \{\mathbf{w} : b(\mathbf{v},\mathbf{w}) = 1\}$ for non-zero $\mathbf{v}$. It is readily checked that these fourteen sets form a 3-design. Perform this construction with each point designated as the origin to obtain a partition. This one is invariant under the group $\mathrm{ASL}(2,3)$ generated by the translations and $\mathrm{Sp}(2,3) = \mathrm{SL}(2,3)$, of order $9 \cdot 8 \cdot 3 = 216$, and there are $9!/216 = 1680$ partitions isomorphic to this one.

Now we show that there are no others. We use the terminology of coding theory. Note that the fourteen words of weight 4 supporting planes of $\mathrm{AG}(3,2)$, together with the all-0 and all-1 words, form the extended Hamming code of length 8 (the code we met in Section 3.2, extended by an overall parity check); it is the only *doubly-even self-dual code* of length 8 (that is, the only code $C = C^\perp$ with all weights divisible by 4).

Let V be the vector space of all words of length 9 and even weight. The function $f(\mathbf{v}) = \frac{1}{2}\mathrm{wt}(\mathbf{v}) \pmod 2$ is a quadratic form on V, which polarises to the usual dot product. Thus maximal t.s. subspaces for f are just doubly even self-dual codes, and their existence shows that f has rank 4 and so is the split

form defining the triality quadric. (The quadric Q consists of the words of weight 4 and 8.)

Suppose we have a partition of the 4-sets into nine affine spaces. An easy counting argument shows that every point is excluded by just one of the designs. So if we associate with each design the word of weight 8 whose support is its point set, we obtain a solid on the quadric, and indeed a *spread* or partition of the quadric into solids.

All these solids belong to the same family, since they are pairwise disjoint. So we can apply the triality map and obtain a set of nine points which are pairwise non-collinear, that is, an *ovoid*. Conversely, any ovoid gives a spread. In fact, an ovoid gives a spread of solids of each family, by applying triality and its inverse. So the total number of spreads is twice the number of ovoids.

The nine words of weight 8 form an ovoid. Any ovoid is equivalent to this one. (Consider the Gram matrix of inner products of the vectors of an ovoid; this must have zeros on the diagonal and ones elsewhere.) The stabiliser of this ovoid is the symmetric group S_9. So the number of ovoids is the index of S_9 in the orthogonal group, which turns out to be 960. Thus, the total number of spreads is $1920 = 240 + 1680$, and we have them all! ■

8.8 Generalised polygons

Projective and polar spaces are important members of a larger class of geometries called *buildings*. Much of the importance of these derives from the fact that they are the "natural" geometries for arbitrary groups of Lie type, just as projective spaces are for linear groups and polar spaces for classical groups. The groups of Lie type include, in particular, all the non-abelian finite simple groups except for the alternating groups and the twenty-six sporadic groups. I do not intend to discuss buildings here — for this, see the lecture notes of Tits [S] or the recent books by Brown [C] and Ronan [P] — but will consider the rank 2 buildings, or *generalised polygons* as they are commonly known. These include the 2-dimensional projective and polar spaces (that is, projective planes and generalised quadrangles).

Recall that a rank 2 geometry has two types of varieties, with a symmetric incidence relation; it can be thought of as a bipartite graph. We use graph-theoretic terminology in the following definition. A rank 2 geometry is a *generalised n-gon* (where $n \geq 2$) if

(GP1) it is connected with diameter n and girth $2n$;

(GP2) for any variety x, there is a variety y at distance n from x.

It is left to the reader to check that, for $n = 2,3,4$, this definition coincides with that of a digon, generalised projective plane or generalised quadrangle respectively.

Let $\mathcal{G}$ be a generalised n-gon. The *flag geometry* of $\mathcal{G}$ has as POINTs the varieties of $\mathcal{G}$ (of both types), and as LINEs the flags of $\mathcal{G}$, with the obvious incidence between POINTs and LINEs. It is easily checked to be a generalised $2n$-gon in which every line has two points; and any generalised $2n$-gon with two points per line is the flag geometry of a generalised n-gon. In future, we usually assume that our polygons are *thick*, that is, have at least three varieties of one type incident with each variety of the other type. It is also easy to show that a thick generalised polygon has *orders*, that is, the number of points per line and the number of lines per point are both constant; and, if n is odd, then these two constants are equal. [Hint: in general, if varieties x and y have distance n, then each variety incident with x has distance $n-2$ from a unique variety incident with y, and *vice versa*.]

We let $s+1$ and $t+1$ denote the numbers of points per line or lines per point, respectively, with the proviso that either or both may be infinite. (If both are finite, then the geometry is finite.) The geometry is thick if and only if $s,t > 1$. The major theorem about finite generalised polygons is the *Feit–Higman theorem* (Feit and Higman [17]:

Theorem 8.13 *A thick generalised n-gon can exist only for $n = 2,3,4,6$ or 8.* ■

In the course of the proof, Feit and Higman derive additional information:

- if $n = 6$, then st is a square;
- if $n = 8$, then $2st$ is a square.

Subsequently, further numerical restrictions have been discovered; for example:

- if $n = 4$ or $n = 8$, then $t \leq s^2$ and $s \leq t^2$;
- if $n = 6$, then $t \leq s^3$ and $s \leq t^3$.

In contrast to the situation for $n = 3$ and $n = 4$, the only known finite thick generalised 6-gons and 8-gons arise from groups of Lie type. There are 6-gons

with $s = t = q$ and with $s = q$, $t = q^3$ for any prime power q; and 8-gons with $s = q$, $t = q^2$, where q is an odd power of 2. In the next section, we discuss a class of 6-gons including the first-mentioned finite examples.

There is no hope of classifying infinite generalised n-gons, which exist for all n (Exercise 2). However, assuming a symmetry condition, the *Moufang condition*, which generalises the existence of central collineations in projective planes, and is also equivalent to a generalisation of Desargues' theorem, Tits [35, 36] and Weiss [39] derived the same conclusion as Feit and Higman, namely, that $n = 2$, 3, 4, 6 or 8.

As for quadrangles, the question of the existence of thick generalised n-gons (for $n \geq 3$) with s finite and t infinite is completely open. Of course, n must be even in such a geometry!

Exercises

1. Prove the assertions claimed to be "easy" in the text.
2. Construct infinite "free" generalised n-gons for any $n \geq 3$.

8.9 Some generalised hexagons

In this section, we use triality to construct a generalised hexagon called $G_2(F)$ over any field F. The construction is due to Tits. The name arises from the fact that the automorphism groups of these hexagons are the Chevalley groups of type G_2, as constructed by Chevalley from the simple Lie algebra G_2 over the complex numbers.

We begin with the triality quadric Q. Let $\mathbf{v}$ be a non-singular vector. Then $\mathbf{v}^\perp \cap Q$ is a rank 3 quadric. Its maximal t.s. subspaces are planes, and each lies in a unique solid of each family on Q. Conversely, a solid on Q meets $\mathbf{v}^\perp$ in a plane. Thus, fixing $\mathbf{v}$, there is are bijections between the two families of solids and the set of planes on $Q' = Q \cap \mathbf{v}^\perp$. On this set, we have the structure of the *dual polar space* induced by the quadric Q'; in other words, the POINTs are the planes on this quadric, the LINES are the lines, and incidence is reversed inclusion. Call this geometry $\mathcal{G}$.

Applying triality, we obtain a representation of $\mathcal{G}$ using all the points and some of the lines of Q.

Now we take a non-singular vector, which may as well be the same as the vector $\mathbf{v}$ already used. (Since we have applied triality, there is no connection.)

The geometry $\mathcal{H}$ consists of those points and lines of $\mathcal{G}$ which lie in $\mathbf{v}^\perp$. Thus, it consists of all the points, and some of the lines, of the quadric Q'.

Theorem 8.14 *$\mathcal{H}$ is a generalised hexagon.*

Proof First we observe some properties of the geometry $\mathcal{G}$, whose points and lines correspond to planes and lines on the quadric Q'. The distance between two points is equal to the codimension of their intersection. If two planes of Q' meet non-trivially, then the corresponding solids of Q (in the same family) meet in a line, and so (applying triality) the points are perpendicular. Hence:

(a) Points of $\mathcal{G}$ lie at distance 1 or 2 if and only if they are perpendicular.

Let x, y, z, w be four points of $\mathcal{G}$ forming a 4-cycle. These points are pairwise perpendicular (by (a)), and so they span a t.s. solid S. We prove:

(b) The geometry induced on S by $\mathcal{G}$ is a symplectic GQ.

Keep in mind the following transformations:

solid S
$\to$ point p (by triality)
$\to$ quadric $\overline{Q}$ in $p^\perp/p$ (residue of p)
$\to \mathrm{PG}(3,F)$ (Klein correspondence).

Now points of S become solids of one family containing p, then planes of one family in $\overline{Q}$, then points in $\mathrm{PG}(3,F)$; so we can identify the two ends of this chain.

Lines of $\mathcal{G}$ in S become lines through p perpendicular to $\mathbf{v}$, then points of $\overline{Q}$ perpendicular to $\langle \overline{\mathbf{v}} \rangle = \langle \mathbf{v}, p \rangle / p$, then t.i. lines of a symplectic GQ, by the correspondence described in Section 8.3. Thus (b) is proved.

A property of $\mathcal{G}$ established in Proposition 7.9 is:

(c) If x is a point and L a line, then there is a unique point of L nearest to x.

We now turn our attention to $\mathcal{H}$, and observe first:

(d) Distances in $\mathcal{H}$ are the same as in $\mathcal{G}$.

For clearly distances in $\mathcal{H}$ are at least as great as those in $\mathcal{G}$, and two points of $\mathcal{H}$ at distance 1 (i.e., collinear) in $\mathcal{G}$ are collinear in $\mathcal{H}$.

Suppose that $x, y \in \mathcal{H}$ lie at distance 2 in $\mathcal{G}$. They are joined by more than one path of length 2 there, hence lie in a solid S carrying a symplectic GQ, as

in (b). The points of $\mathcal{H}$ in S are those of $S \cap \mathbf{v}^\perp$, a plane on which the induced substructure is a plane pencil of lines of $\mathcal{H}$. Hence x and y lie at distance 2 in $\mathcal{H}$.

Finally, let $x, y \in \mathcal{H}$ lie at distance 3 in $\mathcal{G}$. Take a line L of $\mathcal{H}$ through y; there is a point z of $\mathcal{G}$ (and hence of $\mathcal{H}$) on L at distance 2 from x (by (c)). So x and y lie at distance 3 in $\mathcal{H}$.

In particular, property (c) holds also in $\mathcal{H}$.

(e) For any point x of $\mathcal{H}$, the lines of $\mathcal{H}$ through x form a plane pencil.

For, by (a), the union of these lines lies in a t.s. subspace, hence they are coplanar; there are no triangles (by (c)), so this plane contains two points at distance 2; now the argument for (d) applies.

Finally:

(f) $\mathcal{H}$ is a generalised hexagon.

We know it has diameter 3, and (GP2) is clearly true. A circuit of length less than 6 would be contained in a t.s. subspace, leading to a contradiction as in (d) and (e). (In fact, by (c), it is enough to exclude quadrangles.) ■

Cameron and Kantor [12] give a more elementary construction of this hexagon. Their construction, while producing the embedding in Q', depends only on properties of the group $\mathrm{PSL}(3,F)$. However, the proof that it works uses both counting arguments and arguments about finite groups; it is not obvious that it works in general, although the result remains true.

If F is a perfect field of characteristic 2 then, by Theorem 8.5, Q' is isomorphic to the symplectic polar space of rank 3; so $\mathcal{H}$ is embedded as all the points and some of the lines of $\mathrm{PG}(5,F)$.

Two further results will be mentioned without proof. First, if the field F has an automorphism of order 3, then the construction of $\mathcal{H}$ can be "twisted", much as can be done to the Klein correspondence to obtain the duality between orthogonal and unitary quadrangles (mentioned in Section 8.3), to produce another generalised hexagon, called ${}^3D_4(F)$. In the finite case, ${}^3D_4(q^3)$ has parameters $s = q^3$, $t = q$.

Second, there is a construction similar to that of Section 8.4. The generalised hexagon $G_2(F)$ is self-dual if F is a perfect field of characteristic 3, and is self-polar if F has an automorphism σ satisfying $\sigma^2 = 3$. In this case, the set of absolute points of the polarity is an ovoid, a set of pairwise non-collinear points meeting every line of $\mathcal{H}$, and the group of collineations commuting with the polarity has as a normal subgroup the *Ree group* ${}^2G_2(F)$, acting 2-transitively on the points of the ovoid.

Exercise

1. Show that the hexagon $\mathcal{H}$ has two disjoint planes E and F, each of which consists of pairwise non-collinear (but perpendicular) points. Show that each point of E is collinear (in $\mathcal{H}$) to the points of a line of F, and dually, so that E and F are naturally dual. Show that the points of $E \cup F$, and the lines of $\mathcal{H}$ joining their points, form a non-thick generalised hexagon which is the flag geometry of $\mathrm{PG}(2, F)$. (This is the starting point in the construction of Cameron and Kantor referred to in the text.)

9

The geometry of the Mathieu groups

The topic of this chapter is something of a diversion, but is included for two reasons: first, its intrinsic interest; and second, because the geometries described here satisfy axioms not too different from those we have seen for projective, affine and polar spaces, and so they indicate the natural boundaries of the theory.

9.1 The Golay code

The basic concepts of coding theory were introduced in Section 3.2, where we also saw that a non-trivial perfect 3-error-correcting code must have length 23 (see Exercise 3.2.2). Such a code C may be assumed to contain the zero word (by translation), and so any other word has weight at least 7; and

$$|C| = \frac{2^{23}}{\binom{23}{0} + \binom{23}{1} + \binom{23}{2} + \binom{23}{3}} = 2^{12}.$$

We extend C to a code $\overline{C}$ of length 24 by adding an *overall parity check*; that is, we put a 0 in the 24^{th} coordinate of a word whose weight (in C) is even, and a 1 in a word whose weight is odd. The resulting code has all words of even weight, and hence all distances between words even; since adding a coordinate cannot decrease the distance between words, the resulting code has minimum distance 8.

In this section, we outline a proof of the following result.

Theorem 9.1 *There is a unique code with length* 24*, minimum distance* 8*, and containing* 2^{12} *codewords one of which is zero (up to coordinate permutations).* ∎

This code is known as the *(extended binary) Golay code.* It is a linear code (the linearity does not have to be assumed).

Remark There are many constructions of this code; for an account of some of these, see Cameron and Van Lint [F]. As a general principle, a good construction of an object leads to a proof of its uniqueness (by showing that it must be constructed this way), thence to a calculation of its automorphism group (since the object is uniquely built around a starting configuration, and so any isomorphism between such starting configurations extends uniquely to an automorphism), and gives on the way a subgroup of the automorphism group (consisting of the automorphism group of the starting configuration). This point will not be laboured below, but the interested reader may like to examine this and other constructions from this point of view. The particular construction given here has been chosen for two reasons: first, as an application of the Klein correspondence; and second, since it makes certain properties of the automorphism group more accessible.

Proof First, we review the isomorphism between $\mathrm{PSL}(4,2)$ and A_8 outlined in Exercise 8.1.1. Let U be the binary vector space consisting of words of even weight and length 8, Z the subspace consisting of the all-zero and all-one words, and $V = U/Z$. The function mapping a word of U to 0 or 1 according as its weight is congruent to 0 or 2 mod 4 induces a quadratic form f on V, whose zeros form the Klein quadric Q; let W be the vector space of rank 4 whose lines are bijective with the points of Q. Note that the points of Q correspond to partitions of $N = \{1,\ldots,8\}$ into two subsets of size 4.

Let $\Omega = N \cup W$. This set will index the coordinates of the code C we construct. A words of C will be specified by its support, a subset of N and a subset of W. In particular, $\emptyset, N, W$ and $N \cup W$ will be words; so we can complement the subset of N or the subset of W defining a word and obtain another word.

The first non-trivial class of words is obtained by combining the empty subset of N (or the whole of N) with any hyperplane in W (or its coset).

A complementary pair of 4-subsets of N corresponds to a point of Q, and hence to a line L in W. Each 4-subset of N, together with any coset of the corresponding L, is a codeword. Further words are obtained by replacing the coset of L by its symmetric difference with a coset of a hyperplane not containing L (such a coset meets L in two vectors).

A 2-subset of N, or the complementary 6-subset, represents a non-singular point, which translates into a symplectic form b on W. The quadric associated with any quadratic form which polarises to b, together with the 2-subset of N, defines a codeword.

This gives us a total of

$$4+4\cdot 15+\binom{8}{4}\cdot(4+4\cdot 7)+\binom{8}{2}\cdot 16\cdot 4=2^{12}$$

codewords. Moreover, a fairly small amount of case checking shows that the code is linear. Its minimum weight is visibly 8.

We now outline the proof that there is a unique code C of length 24, cardinality 2^{12}, and minimum weight 8, containing $\mathbf{0}$. Counting arguments show that such a code contains 759 words of weight 8, 2576 of weight 12, 759 of weight 16, and the all-1 word $\mathbf{1}$ of weight 24. Now, if the code is translated by any codeword, the hypotheses still hold, and so the conclusion about weights does too. Thus, the distances between pairs of codewords are 0, 8, 12, 16, and 24. It follows that all inner products are zero, so $C\subset C^{\perp}$; it then follows from the cardinality that $C=C^{\perp}$, and in particular C is a linear code.

Let N be an octad, and W its complement. Restriction of codewords to N gives a homomorphism θ from C to a code of length 8 in which all words have even weight. It is readily checked that every word of even weight actually occurs. So the kernel of θ has rank 5. This kernel is a code of length 16 and minimum weight 8. There is a unique code with these properties: it consists of the all-zero and all-one words, together with the characteristic functions of hyperplanes of a rank 4 vector space. (This is the *first-order Reed–Muller code* of length 16.) Thus we have identified W with a vector space, and found the first non-trivial class of words in the earlier construction.

Now, to be brief: if B is an octad meeting N in four points, then $B\cap W$ is a line; if $|B\cap N|=2$, then $B\cap W$ is a quadric; and all the other details can be checked, given sufficient perseverence. ■

The automorphism group of the extended Golay code is the 54-transitive *Mathieu group* M_{24}. This is one of only two finite 5-transitive groups other than symmetric and alternating groups; it is one of the first of the 26 "sporadic" simple groups to be found; and its geometry is the starting point for the construction of many other sporadic groups (the Conway and Fischer groups and the "Monster"). The group M_{24} will be considered further in Section 9.4.

9.2 The Witt system

Let X be the set of coordinate positions of the Golay code G. Now any word can be identified uniquely with the subset of X consisting of the positions where

it has entries equal to 1 (its *support*). Let $\mathcal{B}$ be the set of supports of the 759 codewords of weight 8. An element of $\mathcal{B}$ is called an *octad*; the support of a word of weight 12 in G is called a *dodecad*.

From the linearity of G, we see that the symmetric difference of two octads is the support of a word of G, necessarily an octad, a dodecad, or the complement of an octad; the intersection of the two octads has cardinality 4, 2 or 0 respectively. Three pairwise disjoint octads form a *trio*. (In our construction of the extended Golay code in the last section, the three "blocks" of eight coordinates form a trio.)

Proposition 9.2 $(X, \mathcal{B})$ *is a* 5-$(24, 8, 1)$ *design or Steiner system.*

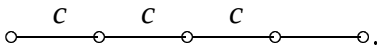
Proof As we have just seen, it is impossible for two octads to have more than four points in common, so five points lie in at most one octad. Since there are 759 octads, the average number containing five points is $759 \cdot \binom{8}{5} / \binom{24}{5} = 1$; so five points lie in exactly one octad. However, the proposition follows more directly from the properties of the code G.

Take any five coordinates, and delete one of them. The remaining coordinates support a word $\mathbf{v}$ of weight 4. But the Golay code obtained by deleting a coordinate from G is perfect 3-error-correcting, and so contains a unique word $\mathbf{c}$ at distance 3 or less from $\mathbf{v}$. It must hold that $\mathbf{c}$ has weight 7 and its support contains that of $\mathbf{v}$ (and $\mathbf{c}$ is the unique such word). Re-introducing the deleted coordinate (which acts as a parity check for the Golay code), we obtain a unique octad containing the given 5-set. ■

This design is known as the *Witt system*; Witt constructed it from its automorphism group, the Mathieu group M_{24}, though nowdays the procedure is normally reversed.

Now choose any three coordinates, and call them ∞_1, ∞_2, ∞_3. Let $X' = X \setminus \{\infty_1, \infty_2, \infty_3\}$, and let $\mathcal{B}'$ be the set of octads containing the chosen points, with these points removed. Then $(X', \mathcal{B}')$ is a 2-(21, 5, 1) design, that is, a projective plane of order 4. Since there is a unique projective plane of order 4 (see Exercise 4.3.6), it is isomorphic to $\mathrm{PG}(2, 4)$.

Proposition 9.3 *The geometry whose varieties are all subsets of X of cardinalities* 1, 2, 3 *and* 4*, and all octads, with incidence defined by inclusion, belongs to the diagram*



■

The remaining octads can be identified with geometric configurations in $\mathrm{PG}(2,4)$. We outline this, omitting detailed verification. In fact, the procedure can be reversed, and the Witt system constructed from objects in $\mathrm{PG}(2,4)$. See Lüneburg [N] for the details of this construction.

1. An octad containing two of the three points ∞_i corresponds to a set of six points of $\mathrm{PG}(2,4)$ meeting any line in 0 or 2 points, in other words, a hyperoval. All 168 hyperovals occur in this way. If we call two hyperovals "equivalent" if their intersection has even cardinality, we obtain a partition into three classes of size 56, corresponding to the three possible pairs of points ∞_i; so this partition can be defined internally.

2. An octad containing one point ∞_i corresponds to a set of seven points of $\mathrm{PG}(2,4)$ meeting every line in 1 or 3 points, that is, a Baer subplane (when equipped with the lines meeting it in three points). Again, all 360 Baer subplanes occur, and the partition can be intrinsically defined.

3. An octad containing none of the points ∞_i is a set of eight points of $\mathrm{PG}(2,4)$ which is the symmetric difference of two lines. Every symmetric difference of two lines occurs (there are 210 such sets).

Since octads and dodecads also intersect evenly, we can extend this analysis to dodecads. Consider a dodecad containing ∞_1, ∞_2 and ∞_3. It contains nine points of $\mathrm{PG}(2,4)$, meeting every line in 1 or 3 points. These nine points form a unital, the set of absolute points of a unitary polarity (or the set of zeros of a non-degenerate Hermitian form). Their intersections of size 3 with lines form a 2-$(9,3,1)$ design, a Steiner triple system which is isomorphic to $\mathrm{AG}(2,3)$, and is also famous as the *Hessian configuration* of inflection points of a non-singular cubic. (Since the field automorphism of $\mathrm{GF}(4)$ is $\alpha \mapsto \alpha^2$, the Hermitian form $x_0x_1^\alpha + x_1x_0^\alpha + x_2x_2^\alpha$ is a cubic form, and its zeros form a cubic curve; in this special case, every point is an inflection.)

Exercises

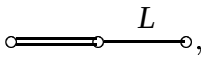
1. Verify the connections between octads and dodecads and configurations in $\mathrm{PG}(2,4)$ claimed in the text.

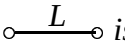
2. Let B be an octad, and $Y = X \setminus B$. Consider the geometry $\mathcal{G}$ whose points are those of Y; whose lines are all pairs of points; whose planes are all sets $B' \setminus B$, where B' is an octad meeting B in four points; and whose solids are the octads disjoint from B. prove that $\mathcal{G}$ is the affine geometry $\mathrm{AG}(4,2)$.

9.3 Sextets

A *tetrad* is a set of four points of the Witt system. Any tetrad is contained in five octads, which partition the remaining twenty points into five tetrads. Now the symmetric difference of two octads intersecting in a tetrad is an octad; so the union of any two of our six tetrads is an octad. A set of six pairwise disjoint tetrads with this property is called a *idxsextet*.

Proposition 9.4 *Let $\mathcal{G}$ be the geometry whose POINTS, LINES and PLANES are the octads, trios and sextets respectively, with incidence defined as follows: a LINE is incident with any POINT it contains; a PLANE is incident with a POINT which is the union of two of its tetrads; and a PLANE is incident with a LINE if it is incident with each POINT of the LINE. Then $\mathcal{G}$ belongs to the diagram*



where  *is the linear space consisting of points and lines of* $\mathrm{PG}(3,2)$.

Proof Calculate residues. Take first a PLANE or sextet. It contains six tetrads; the union of any two of them is a POINT, and any partition into three sets of two is a LINE. This is a representation of the unique GQ with $s = t = 2$ that we saw in Section 7.1.

Now consider the residue of a POINT or octad. We saw in Exercise 9.2.2 that the complement of an octad carries an affine space $\mathrm{AG}(4,2)$; LINEs incident with the POINT correspond to parallel classes of planes in the affine space, and PLANEs incident with it to parallel classes of LINEs. Projectivising and dualising, we see the points and lines of $\mathrm{PG}(3,2)$.

Finally, any POINT and PLANE incident with a common LINE are incident with one another. ■

The geometry does not contain objects which would correspond to the planes of $\mathrm{PG}(3,2)$ in the residue of a point. The diagram is sometimes drawn with a "ghost node" corresponding to these non-existent varieties.

Exercise

1. In the geometry $\mathcal{G}$ of Proposition 9.4, define the distance between two points to be the number of lines on a shortest path joining them. Prove that, if x is a point and L a line, then there is a unique point of L at minimum distance from x.

9.4 The large Mathieu groups

Just as every good construction of the Golay code or the Witt system contains the seeds of a uniqueness proof (as we observed in Section 9.1), so every good uniqueness proof contains the seeds of an argument establishing various properties of its automorphism group (in particular, its order, and some large subgroup, the particular subgroup depending on the construction used). I will outline this for the construction of Section 9.1.

Theorem 9.5 *The automorphism group of the Golay code, or of the Witt system, is a* 5*-transitive simple group of order* $24 \cdot 23 \cdot 22 \cdot 21 \cdot 20 \cdot 48$.

Remark This group is of course the Mathieu group M_{24}. Part of the reason for the construction we gave (not the simplest available!) is that it makes our job now easier.

Proof First note that the design and the code have the same automorphism group; for the code is spanned by the design, and the design is the set of words of weight 8 in the code.

The uniqueness proof shows that the automorphism group is transitive on octads. For, given two copies of the Golay code, and an octad in each, there is an isomorphism between the two codes mapping the chosen octad in the first to that in the second. Also, the stabiliser of an octad preserves the affine space structure on its complement, and (from the construction) induces $\mathrm{AGL}(4,2)$ on it. (It induces A_8 on the octad, the kernel of this action being the translation group of the affine space.) This gives the order of the group.

Given two 5-tuples of distinct points, each lies in a unique octad. There is an automorphism carrying the first octad to the second; then, since A_8 is 5-transitive, we can fix the second octad and map the 5-tuple to the correct place. The 5-transitivity follows.

We also have a subgroup $H = \mathrm{AGL}(4,2)$ of our unknown group G, and it is easily seen that H is maximal. Suppose that N is a non-trivial normal subgroup of G. Then $HN = G$, and $H \cap N$ is a normal subgroup of H, necessarily the identity or the translation group. (If $H \cap N = H$ then $N = G$.) This gives two possibilities for the order of N, namely 759 and $759 \cdot 16$. But N, a normal subgroup of a 5-transitive group, is at least 4-transitive, by an old theorem of Jordan; so $24 \cdot 23 \cdot 22 \cdot 21$ divides $|N|$, a contradiction. We conclude that G is simple. ■

The stabiliser of three points is a group of collineations of $\mathrm{PG}(2,4)$, necessarily $\mathrm{PSL}(3,4)$ (by considering order). The ovals and Baer subplanes each fall into three orbits for $\mathrm{PSL}(3,4)$, these orbits being the classes used in Lüneburg's construction. The set-wise stabiliser of three points is $\mathrm{P\Gamma L}(3,4)$. Looked at another way, Lüneburg's construction and uniqueness proof gives us the subgroup $\mathrm{P\Gamma L}(3,4)$ of M_{24}.

9.5 The small Mathieu groups

To conclude this chapter, I describe briefly the geometry associated with the Mathieu group M_{12}.

There are two quite different approaches. One locates the geometry within the Golay code. The group M_{12} can be defined as the stabiliser of a dodecad in M_{24}; it acts sharply 5-transitively on this dodecad, and on the complementary dodecad, but the two permutation representations are not equivalent. The dodecad D carries a design, which can be seen as follows. It intersects any octad in an even number, at most 6, of points; and any five points of D lie in a unique octad, meeting D in 6 points. So the intersections of size 6 of octads with D are the blocks of a 5-$(12,6,1)$ design or Steiner system.

Alternatively, there are "characteristic 3" objects with properties resembling the binary Golay code. There is a *ternary Golay code*, a set of ternary words of length 12 (that is, entries in $\mathrm{GF}(3)$) forming a subspace of $\mathrm{GF}(3)^{12}$ of rank 6, and having minimum weight 6; the supports of weight 6 of codewords form the blocks of the design. Alternatively, there is a set of 12 points in $\mathrm{PG}(5,3)$ on which M_{12} is induced, as follows. There is a *Hadamard matrix* H of size 12×12 (a matrix with entries ± 1 satisfying $HH^\top = 12I$), unique up to row and column permutations and sign changes; over $\mathrm{GF}(3)$, it has rank 6, and its rows span the required points. Now the design is obtained as follows. The point set is identified with the set of rows. Any two columns agree in six rows and disagree in the other six, defining two sets of size 6 which are blocks of the design; and all $2\cdot\binom{12}{2} = 132$ blocks are obtained in this way.

Some connection between characteristics 2 and 3 can be seen from the observation we made in Section 9.2, that a unital in $\mathrm{PG}(2,4)$ is isomorphic to the affine plane $\mathrm{AG}(2,3)$. It turns out that the three times extensions of these two planes are associated with codes in characteristics 2 and 3 respectively, and that one extension contains the other. However, the large Witt system is not embeddable in $\mathrm{PG}(5,4)$, so the analogy is not perfect.

Exercise

1. Let $G = \mathrm{AG}(2,3)$, and X the set of lines of G (so that $|X| = 12$). Consider the subsets of X of the following types:

- all unions of two parallel classes;
- the lines of two classes containing a point p, and those of the other two not containing p;
- a parallel class, with the lines of the others containing a fixed point p; and the complements of these.

Show that these $6 + 54 + 2 \cdot 36 = 132$ sets of size 6 form a 5-$(12,6,1)$ design. Assuming the uniqueness of this design, prove that $\mathrm{AGL}(2,3) \subseteq M_{12}$.

10

Exterior powers and Clifford algebras

In this chapter, various algebraic constructions (exterior products and Clifford algebras) are used to embed some geometries related to projective and polar spaces (subspace and spinor geometries) into projective spaces. In the process, we learn more about the geometries themselves.

10.1 Tensor and exterior products

Throughout this chapter, F is a commutative field (except for a brief discussion of why this assumption is necessary).

The *tensor product* $V \otimes W$ of two F-vector spaces V and W is the free-est bilinear product of V and W: that is, if (as customary), we write the product of vectors $\mathbf{v} \in V$ and $\mathbf{w} \in W$ as $\mathbf{v} \otimes \mathbf{w}$, then we have

$$(\mathbf{v}_1 + \mathbf{v}_2) \otimes \mathbf{w} = \mathbf{v}_1 \otimes \mathbf{w} + \mathbf{v}_2 \otimes \mathbf{w}, \qquad (\alpha\mathbf{v}) \otimes \mathbf{w} = \alpha(\mathbf{v} \otimes \mathbf{w}),$$
$$\mathbf{v} \otimes (\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v} \otimes \mathbf{w}_1 + \mathbf{v} \otimes \mathbf{w}_2, \qquad \mathbf{v} \otimes (\alpha\mathbf{w}) = \alpha(\mathbf{v} \otimes \mathbf{w}).$$

Formally, we let X be the F-vector space with basis consisting of all the ordered pairs $(\mathbf{v}, \mathbf{w})$ ($\mathbf{v} \in V, \mathbf{w} \in W$), and Y the subspace spanned by all expressions of the form $(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) - (\mathbf{v}_1, \mathbf{w}) - (\mathbf{v}_2, \mathbf{w})$ and three similar expressions; then $V \otimes W = X/Y$, with $\mathbf{v} \otimes \mathbf{w}$ the image of $(\mathbf{v}, \mathbf{w})$ under the canonical projection. Sometimes, to emphasize the field, we write $V \otimes_F W$.

This construction will only work as intended over a commutative field. For

$$\alpha\beta(\mathbf{v} \otimes \mathbf{w}) = \alpha(\beta\mathbf{v} \otimes \mathbf{w}) = \beta\mathbf{v} \otimes \alpha\mathbf{w} = \beta(\mathbf{v} \otimes \alpha\mathbf{w}) = \beta\alpha(\mathbf{v} \otimes \mathbf{w}),$$

so if $\mathbf{v}\otimes\mathbf{w}\neq 0$ then $\alpha\beta=\beta\alpha$.

There are two representations convenient for calculation. If V has a basis $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ and W a basis $\{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$, then $V\otimes W$ has a basis

$$\{\mathbf{v}_i\otimes\mathbf{w}_j : 1\leq i\leq n, 1\leq j\leq m\}.$$

If V and W are identified with F^n and F^m respectively, then $V\otimes W$ can be identified with the space of $n\times m$ matrices over F, where $\mathbf{v}\otimes\mathbf{w}$ is mapped to the matrix $\mathbf{v}^\top\mathbf{w}$.

In particular, $\mathrm{rk}(V\otimes W)=\mathrm{rk}(V)\cdot\mathrm{rk}(W)$.

Suppose that V and W are F-algebras (that is, have an associative multiplication which is compatible with the vector space structure). Then $V\otimes W$ is an algebra, with the rule

$$(\mathbf{v}_1\otimes\mathbf{w}_1)\cdot(\mathbf{v}_2\otimes\mathbf{w}_2)=(\mathbf{v}_1\cdot\mathbf{v}_2)\otimes(\mathbf{w}_1\cdot\mathbf{w}_2).$$

Of course, we can form the tensor product of a space with itself; and we can form iterated tensor products of more than two spaces. Let $\bigotimes^k V$ denote the k-fold tensor power of V. Now the *tensor algebra* of V is defined to be

$$T(V)=\bigoplus_{k=0}^{\infty}(\bigotimes^k V),$$

with multiplication given by the rule

$$(\mathbf{v}_1\otimes\ldots\otimes\mathbf{v}_n)\cdot(\mathbf{v}_{n+1}\otimes\ldots\otimes\mathbf{v}_{m+n})=\mathbf{v}_1\otimes\ldots\otimes\mathbf{v}_{m+n}$$

on homogeneous elements, and extended linearly. It is the free-est associative algebra generated by V.

The *exterior square* of a vector space V is the free-est bilinear square of V in which the square of any element of V is zero. In other words, it is the quotient of $\bigotimes^2 V$ by the subspace generated by all vectors $\mathbf{v}\otimes\mathbf{v}$ for $\mathbf{v}\in V$. We write it as $\bigwedge^2 V$, or $V\wedge V$, and denote the product of $\mathbf{v}$ and $\mathbf{w}$ by $\mathbf{v}\wedge\mathbf{w}$. Note that $\mathbf{w}\wedge\mathbf{v}=-\mathbf{v}\wedge\mathbf{w}$. If $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ is a basis for V, then a basis for $V\wedge V$ consists of all vectors $\mathbf{v}_1\wedge\mathbf{v}_j$, for $1\leq i<j\leq n$; so

$$\mathrm{rk}(V\wedge V)=\binom{n}{2}=\tfrac{1}{2}n(n-1).$$

More generally, we can define the k^{th} *exterior power* $\bigwedge^k V$ as a k-fold multilinear product, in which any product of vectors vanishes if two factors are equal.

Its basis consists of all expressions $\mathbf{v}_{i_1} \wedge \ldots \wedge \mathbf{v}_{i_k}$, with $1 \leq i_1 < \ldots < i_k \leq n$; and its dimension is $\binom{n}{k}$. Note that $\bigwedge^k V = 0$ if $k > n = \mathrm{rk}(V)$.

The *exterior algebra* of V is

$$\bigwedge(V) = \bigoplus_{k=0}^{n} (\bigwedge^k V),$$

with multiplication defined as for the tensor algebra. Its rank is $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

If θ is a linear transformation on V, then θ induces in a natural way linear transformations $\bigotimes^k \theta$ on $\bigotimes^k V$, and $\bigwedge^k \theta$ on $\bigwedge^k V$, for all k. If $\mathrm{rk}(V) = n$, then we have $\mathrm{rk}(\bigwedge^n V) = 1$, and so $\bigwedge^n \theta$ is a scalar. In fact, $\bigwedge^n \theta = \det(\theta)$. (This fact is the basis of an abstract, matrix-free, definition of the determinant.)

Exercises

1. Let F be a skew field, V a right F-vector space, and W a left vector space. Show that it is possible to define $V \otimes_F W$ as an abelian group so that

$$(\mathbf{v}_1 + \mathbf{v}_2) \otimes \mathbf{w} = \mathbf{v}_1 \otimes \mathbf{w} + \mathbf{v}_2 \otimes \mathbf{w}, \quad \mathbf{v} \otimes (\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v} \otimes \mathbf{w}_1 + \mathbf{v} \otimes \mathbf{w}_2$$

and

$$(\mathbf{v}\alpha) \otimes \mathbf{w} = \mathbf{v} \otimes (\alpha \mathbf{w}).$$

2. In the identification of $F^n \otimes F^m$ with the space of $n \times m$ matrices, show that the rank of a matrix is equal to the minimum r for which the corresponding tensor can be expressed in the form $\sum_{i=1}^{r} \mathbf{v}_i \otimes \mathbf{w}_i$. Show that, in such a minimal expression, $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent, as are $\mathbf{w}_1, \ldots, \mathbf{w}_r$.

3. (a) If K is an extension field of F, and n a positive integer, prove that

$$M_n(F) \otimes_F K \cong M_n(K),$$

where $M_n(F)$ is the ring of $n \times n$ matrices over F.

(b) Prove that $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C} \oplus \mathbb{C}$.

4. Define the *symmetric square* S^2V of a vector space V, the free-est bilinear square of V in which $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$. Find a basis for it, and calculate its dimension. More generally, define the k^{th} *symmetric power* S^kV, and calculate its dimension; and define the *symmetric algebra* $S(V)$. If $\dim(V) = n$, show that the symmetric algebra on V is isomorphic to the polynomial ring in n variables over the base field.

5. Prove that, if θ is a linear map on V, where $\mathrm{rk}(V) = n$, then $\bigwedge^n \theta = \det(\theta)$.

10.2 The geometry of exterior powers

Let V be an F-vector space of rank n, and k a positive integer less than n. There are a couple of ways of defining a geometry on the set $\Sigma_k = \Sigma_k(V)$ of subspaces of V of rank k (equivalently, the $(k-1)$-dimensional subspaces of $\mathrm{PG}(\binom{n}{k}-1, F)$, which I now describe.

The first approach produces a point-line geometry. For each pair U_1, U_2 of subspaces of V with $U_1 \subset U_2$, $\mathrm{rk}(U_1) = k-1$, $\mathrm{rk}(U_2) = k+1$, a *line*

$$L(U_1, U_2) = \{W \in \Sigma_k : U_1 \subset W \subset U_2\}.$$

Now two points lie in at most one line. For, if W_1, W_2 are distinct subspaces of rank k and $W_1, W_2 \in L(U_1, U_2)$, then $U_1 \subseteq W_1 \cap W_2$ and $\langle W_1, W_2\rangle \subseteq U_2$; so equality must hold in both places. Note that two subspaces are collinear if and only if their intersection has codimension 1 in each. We call this geometry a *subspace geometry*.

In the case $k = 2$, the points of the subspace geometry are the lines of $\mathrm{PG}(n-1, F)$, and its lines are the plane pencils. In particular, for $k = 2$, $n = 4$, it is the Klein quadric.

The subspace geometry has the following important property:

Proposition 10.1 *If three points are pairwise collinear, then they are contained in a projective plane. In particular, a point not on a line L is collinear with none, one or all points of L.*

Proof Clearly the second assertion follows from the first. In order to prove the first assertion, note that there are two kinds of projective planes in the geometry, consisting of all points W (i.e., subspaces of rank k) satisfying $U_1 \subset W \subset U_2$, where either $\mathrm{rk}(U_1) = k-1$, $\mathrm{rk}(U_2) = k+2$, or $\mathrm{rk}(U_1) = k-2$, $\mathrm{rk}(U_2) = k+1$.

So let W_1, W_2, W_3 be pairwise collinear points. If $\mathrm{rk}(W_1 \cap W_2 \cap W_3) = k-1$, then the three points are contained in a plane of the first type; so suppose not. Then we have $\mathrm{rk}(W_1 \cap W_2 \cap W_3) = k-2$; and, by factoring out this intersection, we may assume that $k = 2$. In the projective space, W_1, W_2, W_3 are now three pairwise intersecting lines, and so are coplanar. Thus $\mathrm{rk}\langle W_1, W_2, W_3\rangle = k+1$, and our three points lie in a plane of the second type. ∎

A point-line geometry satisfying the second conclusion of Proposition 10.1 is called a *gamma space*. Gamma spaces are a natural generalisation of polar spaces

(in the Buekenhout–Shult sense), and this property has been used in several recent characterisations (some of which are surveyed by Shult [29]).

The subspace geometries have natural embeddings in projective spaces given by exterior powers, generalising the Klein quadric. Let $X = \bigwedge^k V$; we consider the projective space $\mathrm{PG}(N-1,F)$ based on X, where $N = \binom{n}{k}$. This projective space contains some distinguished points, those spanned by the vectors of the form $\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_k$, for $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$. We call these *pure products*.

Theorem 10.2 *(a)* $\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_k = 0$ *if and only if* $\mathbf{v}_1, \ldots, \mathbf{v}_k$ *are linearly dependent.*

(b) The set of points of $\mathrm{PG}(N-1,F)$ *spanned by non-zero pure products, together with the lines meeting this set in more than two points, is isomorphic to the subspace geometry* $\Sigma_k(V)$.

Proof (a) If $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent, then they form part of a basis, and their product is one of the basis vectors of X, hence non-zero. Conversely, if these vectors are dependent, then one of them can be expressed in terms of the others, and the product is zero (using linearity and the fact that a product with two equal terms is zero).

(b) It follows from our remarks about determinants that, if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are replaced by another k-tuple with the same span, then $\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_k$ is multiplied by a scalar factor, and the point of $\mathrm{PG}(N-1,F)$ it spans is unaltered. If $W_1 \neq W_2$, then we can (as usual in linear algebra) choose a basis for V containing bases for both W_1 and W_2; the corresponding pure products are distinct basis vectors of X, and so span distinct points. The correspondence is one-to-one.

Suppose that W_1 and W_2 are collinear in the subspace geometry. then they have bases $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{w}_1\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{w}_2\}$. Then the points spanned by the vectors

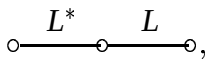
$$\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_{k-1} \wedge (\alpha\mathbf{w}_1 + \beta\mathbf{w}_2)$$

form a line in $\mathrm{PG}(N-1,F)$ and represent all the points of the line in the subspace geometry joining W_1 and W_2.

Conversely, suppose that $\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_k$ and $\mathbf{w}_1 \wedge \ldots \wedge \mathbf{w}_k$ are two pure products. By factoring out the intersection of the corresponding subspaces, we may assume that $\mathbf{v}_1, \ldots, \mathbf{w}_k$ are linearly independent. If $k > 1$, then no other vector in the span of these two pure products is a pure product. If $k = 1$, then the three points are coplanar. ■

The other natural geometry on the set $\Sigma_k(V)$ is just the truncation of the projective geometry to ranks $k-1, k$ and $k+1$; in other words, its varieties are the subspaces of V of these three ranks, and incidence is inclusion. This geometry has no immediate connection with exterior algebra; but it (or the more general form based on any generalised projective geometry) has a beautiful characterisation due to Sprague (1981).

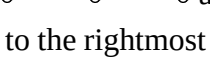
Theorem 10.3 *(a) The geometry just described has diagram*



where L^* denotes the class of dual linear spaces.

(b) Conversely, any geometry with this diagram, in which chains of subspaces are finite, consists of the varieties of ranks $k-1, k$ and $k+1$ of a generalised projective space of finite dimension, two varieties incident if one contains the other.

Proof The residue of a variety of rank $k-1$ is the quotient projective space; and the residue of a variety of rank $k+1$ is the dual of $\mathrm{PG}(k,F)$. This establishes the diagram.

I will not give the proof of Sprague's theorem. the proof is by induction (hence the need to assume finite rank). Sprague shows that it is possible to recognise in the geometry objects corresponding to varieties of rank $k-2$, these objects together with the left and centre nodes forming the diagram 

again, but with the dimension of the residue of a variety belonging to the rightmost node reduced by 1. After finitely many steps, we reach the points, lines and planes of the projective space, which is recognised by the Veblen–Young axioms. ■

Exercise

1. Show that the dual of the generalised hexagon $G_2(F)$ constructed in Section 8.8 is embedded in the subspace geometry of lines of $\mathrm{PG}(6,F)$. [Hint: the lines of the hexagon through a point x are all those containing x in a plane $W(x)$.]

10.3 Near polygons

In this section we consider certain special point-line geometries. These geometries will always be connected, and the *distance* between two points is the

smallest number of lines in a path joining them. A *near polygon* is a geometry with the following property:

(NP) Given any point p and line L, there is a unique point of L nearest to p.

If a near polygon has diameter n, it is called a *near* $2n$*-gon*.

We begin with some elementary properties of near polygons.

Proposition 10.4 *In a near polygon,*

(a) two points lie on at most one line;

(b) the shortest circuit has even length.

Proof (a) Suppose that lines L_1, L_2 contain points p_1, p_2. Let $q \in L_1$. Then q is at distance 1 from the two points p_1, p_2 of L_2, and so is at distance 0 from a unique point of L_2; that is, $q \in L_2$. So $L_1 \subseteq L_2$; and, interchanging these two lines, we find that $L_1 = L_2$.

If a circuit has odd length $2m+1$, then a point lies at distance m from two points of the opposite line; so it lies at distance $m-1$ from some point of this line, and a circuit of length $2m$ is formed. ■

Any generalised polygon is a near polygon; and any "non-degenerate" near 4-gon is a generalised quadrangle (see Exercise 1).

Some deeper structural properties are given in the next two theorems, which were found by Shult and Yanushka [30].

Theorem 10.5 *Suppose that $x_1x_2x_3x_4$ is a circuit of length 4 in a near polygon, at least one of whose sides contains more than two points. Then there is a unique subspace containing these four points which is a generalised quadrangle.* ■

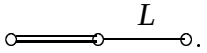
A subspace of the type given by this theorem is called a *quad*.

Corollary 10.6 *Suppose that a near polygon has the properties*

(a) any line contains more than two points;

(b) any two points at distance 2 are contained in a circuit of length 4.

Then the points, lines and quads form a geometry belonging to the diagram



We now assume that the hypotheses of this Corollary apply. Let p be a point and Q a quad. We say that the pair (p,Q) is *classical* if

(a) there is a unique point x of Q nearest p;

(b) for $y \in Q$, $d(y,p) = d(x,p)+1$ if and only if y is collinear with x.

(The point x is the "gateway" to Q from p.) An *ovoid* in a generalised quadrangle is a set O of (pairwise non-collinear) points with the property that any further point of the quadrangle is collinear with a unique point of O. The point-quad pair (p,Q) is *ovoidal* if the set of points of Q nearest to p is an ovoid of Q.

Theorem 10.7 *In a near polygon with at least three points on a line, any point-quad pair is either classical or ovoidal.* ■

A proof in the finite case is outlined in Exercise 2.

We now give an example, the sextet geometry of Section 9.3 (which, as we already know, has the correct diagram). Recall that the POINTs, LINEs, and "QUADs" (as we will now re-name them) of the geometry are the octads, trios and sextets of the Witt system. We check that this is a near polygon, and examine the point-quad pairs.

Two octads intersect in 0, 2 or 4 points. If they are disjoint, they are contained in a trio (i.e., collinear). If they intersect in four points, they define a sextet, and so some octad is disjoint from both; so their distance is 2. If they intersect in two points, their distance is 3. Suppose that $\{B_1,B_2,B_3\}$ is a trio and B an octad not in this trio. Either B is disjoint from (i.e., collinear with) a unique octad in the trio, or its intersections with them have cardinalities 4, 2, 2. In the latter case, it lies at distance 2 from one POINT of the LINE, and distance 3 from the other two.

Now let B be a POINT (an octad), and S a QUAD (a sextet). The intersections of B with the tetrads of S have the property that any two of them sum to 0, 2, 4 or 8; so they are all congruent mod 2. If the intersections have even parity, they are $4,4,0,0,0,0$ (the POINT lies in the QUAD) or $2,2,2,2,0,0$ (B is disjoint from a unique octad incident with S, and the pair is classical). If they have odd parity,

they are $3,1,1,1,1,1$; then B has distance 2 from the five octads containing the first tetrad, and distance 3 from the others. Note that in the GQ of order $(2,2)$, represented as the pairs from a 6-set, the five pairs containing an element of the 6-set form an ovoid. So (B,S) is ovoidal in this case.

10.3.1 Exercises

1. (a) A near polygon with lines of size 2 is a bipartite graph.

(b) A near 4-gon, in which no point is joined to all others, is a generalised quadrangle.

2. Let Q be a finite GQ with order s,t, where $s > 1$.

(a) Suppose that the point set of Q is partitioned into three subsets A,B,C such that for any line L, the values of $|L\cap A|$, $|L\cap B|$ and $|L\cap C|$ are either $1,s,0$, or $0,1,s$. Prove that A is a singleton, and B the set of points collinear with A.

(b) Suppose that the point set of Q is partitioned into two subsets A and B such that any line contains a unique point of A. Prove that A is an ovoid.

(c) Hence prove (10.3.4) in the finite case.

10.4 Dual polar spaces

We now look at polar spaces "the other way up". That is, given an abstract polar space of polar rank n, we consider the geometry whose POINTs and LINEs are the subspaces of dimension $n-1$ and $n-2$ respectively, incidence being reversed inclusion. (This geometry was introduced in Section 7.4.)

Proposition 10.8 *A dual polar space of rank n is a near $2n$-gon.*

Proof This is implicit in what we proved in Proposition 7.9. ■

Any dual polar space has girth 4, and any circuit of length 4 is contained in a unique quad. Moreover, the point-quad pairs are all classical. Both these assertions are easily checked in the polar space by factoring out the intersection of the subspaces in question.

The converse of this result was proved by Cameron [9]. It is stated here using the notation and ideas (and simplifications) of Shult and Yanushka described in the last section.

Theorem 10.9 *Let $\mathcal{G}$ be a near $2n$-gon. Suppose that*

(a) any 4-circuit is contained in a quad;

(b) any point-quad pair is classical;

(c) chains of subspaces are finite.

Then $\mathcal{G}$ is a dual polar space of rank n.

Proof The ideas behind the proof will be sketched.

Given a point p, the residue of p (that is, the geometry of lines and quads containing p) is a linear space, by hypothesis (a). Using (b), it is possible to show that this linear space satisfies the Veblen–Young axioms, and so is a projective space $\mathcal{P}(p)$ (possibly infinite-dimensional). We may assume that this geometry has dimension greater than 2 (otherwise the next few steps are vacuous).

Now, given points p and q, let $\mathcal{X}(p,q)$ be the set of lines through p (i.e., points of $\mathcal{P}(p)$) which belong to geodesics from p to q (that is, which contain points r with $d(q,r) = d(p,q) - 1$). This set is a subspace of $\mathcal{P}(p)$. Let X be any subspace of $\mathcal{P}(p)$, and let

$$\mathcal{Y}(p,X) = \{q : \mathcal{X}(p,q) \subseteq X\}.$$

It can be shown that $\mathcal{Y}(p,X)$ is a subspace of the geometry, containing all geodesics between any two of its points, and that, if p' is any point of $\mathcal{Y}(p,X)$, then there is a subspace X' of $\mathcal{P}(p')$ such that $\mathcal{Y}(p',X') = \mathcal{Y}(p,X)$.

For the final step, it is shown that the subspaces $\mathcal{Y}(p,X)$, ordered by reverse inclusion, satisfy the axioms (P1)–(P4) of Tits. ■

Remark In the case when any line has more than two points, condition (a) is a consequence of (10.3.2), and (10.3.4) shows that (b) is equivalent to the assertion that no point-quad pairs are ovoidal.

10.5 Clifford algebras and spinors

Spinors provide projective embeddings of some geometries related to dual polar spaces, much as exterior powers do for subspace geometries. But they are somewhat elusive, and we have to construct them via Clifford algebras.

Let V be a vector space over a commutative field F, and f a quadratic form on V; let b be the bilinear form obtained by polarising f. The *Clifford algebra* $C(f)$ of f (or of the pair (V,f)) is the free-est algebra generated by V subject to the

condition that $\mathbf{v}^2 = f(\mathbf{v}) \cdot 1$ for all $\mathbf{v} \in V$. In other words, it is the quotient of the tensor algebra $T(V)$ by the ideal generated by all elements $\mathbf{v}^2 - f(\mathbf{v}) \cdot 1$ for $\mathbf{v} \in V$.

Note that $\mathbf{vw} + \mathbf{wv} = b(\mathbf{v}, \mathbf{w}) \cdot 1$ for $\mathbf{v}, \mathbf{w} \in V$.

The Clifford algebra is a generalisation of the exterior algebra, to which it reduces if f is identically zero. And it has the same dimension:

Proposition 10.10 *Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for V. Then $C(f)$ has a basis consisting of all vectors $\mathbf{v}_{i_1} \cdots \mathbf{v}_{i_k}$, for $0 \le i_1 < \ldots < i_k \le n$; and so* $\mathrm{rk}(C(f)) = 2^n$.

Proof Any product of basis vectors can be rearranged into non-decreasing order, modulo products of smaller numbers of basis vectors, using

$$\mathbf{wv} = \mathbf{vw} - b(\mathbf{v}, \mathbf{w}) \cdot 1.$$

A product with two terms equal can have its length reduced. Now the result follows by multilinearity. ■

In an important special case, we can describe the structure of $C(f)$.

Theorem 10.11 *Let f be a split quadratic form of rank n over F (equivalent to*

$$x_1x_2 + x_3x_4 + \ldots + x_{2n-1}x_{2n}.$$

Then $C(f) \cong M_{2^n}(F)$, the algebra of $2^n \times 2^n$ matrices over F.

Proof It suffices to find a linear map $\theta : V \to M_{2^n}(F)$ satisfying

(a) $\theta(V)$ generates $M_{2^n}(F)$ (as algebra with 1);

(b) $\theta(\mathbf{v})^2 = f(\mathbf{v})I$ for all $\mathbf{v} \in V$.

For if so, then $M_{2^n}(F)$ is a homomorphic image of $C(f)$; comparing dimensions, they are equal.

We use induction on n. For $n = 0$, the result is trivial. Suppose that it is true for n, with a map θ. Let $\tilde{V} = V \perp \langle \mathbf{x}, \mathbf{y} \rangle$, where $f(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda\mu$. Define $\tilde{\theta} : \tilde{V} \to M_{2^{n+1}}(F)$ by

$$\tilde{\theta}(\mathbf{v}) = \begin{pmatrix} \theta(\mathbf{v}) & O \\ O & -\theta(\mathbf{v}) \end{pmatrix}, \quad \mathbf{v} \in V,$$

$$\tilde{\theta}(\mathbf{x}) = \begin{pmatrix} O & I \\ O & O \end{pmatrix}, \qquad \tilde{\theta}(\mathbf{y}) = \begin{pmatrix} O & O \\ I & O \end{pmatrix},$$

extended linearly.

To show generation, let $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) \in M_{2^{n+1}}(F)$ be given. We may assume inductively that A, B, C, D are linear combinations of products of $\theta(\mathbf{v})$, with $\mathbf{v} \in V$. The same combinations of products of $\tilde{\theta}(\mathbf{v})$ have the forms $\tilde{A} = \left(\begin{smallmatrix} A & O \\ O & A^* \end{smallmatrix}\right)$, etc. Now

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \tilde{A}\tilde{\theta}(\mathbf{x})\tilde{\theta}(\mathbf{y}) + \tilde{B}\tilde{\theta}(\mathbf{x}) + \tilde{\theta}(\mathbf{y})\tilde{C} + \tilde{\theta}(\mathbf{y})\tilde{D}\tilde{\theta}(\mathbf{x}).$$

To establish the relations, we note that

$$\tilde{\theta}(\mathbf{v} + \lambda\mathbf{x} + \mu\mathbf{y}) = \begin{pmatrix} \theta(\mathbf{v}) & \lambda I \\ \mu I & -\theta(\mathbf{v}) \end{pmatrix},$$

and the square of the right-hand side is $(f(\mathbf{v}) + \lambda\mu)\left(\begin{smallmatrix} I & O \\ O & I \end{smallmatrix}\right)$, as required. ■

More generally, the argument shows the following.

Theorem 10.12 *If the quadratic form f has rank n and germ f_0, then*

$$C(f) \cong C(f_0) \otimes_F M_{2^n}(F).$$

■

In particular, $C(x_0^2 + x_1x_2 + \ldots + x_{2n-1}x_{2n})$ is the direct sum of two copies of $M_{2^n}(F)$; and, if α is a non- square in F, then

$$C(\alpha x_0^2 + x_1x_2 + \ldots + x_{2n-1}x_{2n}) \cong M_{2^n}(K),$$

where $K = F(\sqrt{\alpha})$.

Looked at more abstractly, Theorem 10.12 says that the Clifford algebra of the split form of rank *n* is isomorphic to the algebra of endomorphisms of a vector space S of rank 2^n. This space is the *spinor space*, and its elements are called *spinors*. Note that the connection between the spinor space and the original vector space is somewhat abstract and tenuous! It is the spinor space which carries the geometrical structures we now investigate.

Exercise

1. Prove that the Clifford algebras of the real quadratic forms $-x^2$ and $-x^2 - y^2$ respectively are isomorphic to the complex numbers and the quaternions. What is the Clifford algebra of $-x^2 - y^2 - z^2$?

10.6 The geometry of spinors

In order to connect spinors to the geometry of the quadratic form, we first need to recognise the points of a vector space within its algebra of endomorphisms.

let V be a vector space, A the algebra of linear transformations of V. Then A is a simple algebra. If U is any subspace of V, then

$$I(U) = \{a \in A : \mathbf{v}a \in U \text{ for all } \mathbf{v} \in V\}$$

is a left ideal in A. Every left ideal is of this form (see Exercise 1). So the projective space based on V is isomorphic to the lattice of left ideals of A. In particular, the minimal left ideals correspond to the points of the projective space. Moreover, if U has rank 1, then $I(U)$ has rank n, and A (acting by left multiplication) induces the algebra of linear transformations of U. In this way, the vector space is "internalised" in the algebra.

Now let V carry a split quadratic form of rank n. If U is a totally singular subspace of rank n, then the elements of U generate a subalgebra isomorphic to the exterior algebra of U. Let $\hat{U}$ denote the product of the vectors in a basis of U. Note that $\hat{U}$ is unchanged, apart from a scalar factor, if a different basis is used. Then $\mathbf{v}\mathbf{u}\hat{U} = 0$ whenever $\mathbf{v} \in V$, $\mathbf{u} \in U$, and $\mathbf{u} \neq 0$; so the left ideal generated by $\hat{U}$ has dimension 2^n (with a basis of the form $\{\mathbf{v}_{i_1} \dots \mathbf{v}_{i_k}\hat{U}\}$, where $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of a complement for U, and $1 \leq i_1 < \dots < i_k \leq n$. Thus, $\hat{U}$ generates a minimal left ideal of $C(f)$. By the preceding paragraph, this ideal corresponds to a point of the projective space $\mathrm{PG}(2^n - 1, F)$ based on the spinor space S.

Summarising, we have a map from the maximal totally singular subspaces of the hyperbolic quadric to a subset of the points of projective spinor space. The elements in the image of this map are called *pure spinors*.

We now state some properties of pure spinors without proof.

Proposition 10.13 *(a) There is a decomposition of the spinor space S into two subspaces S^+, S^-, each of rank 2^{n-1}. Any pure spinor is contained in one of these subspaces.*

(b) Any line of spinor space which contains more than two pure spinors has the form

$$\{\langle \hat{U} \rangle : U \text{ is t.s. with rank } n,\ U \text{ has type } \varepsilon,\ U \supset W\},$$

where W is a t.s. subspace of rank $n-2$, and $\varepsilon = \pm 1$. ■

In (a), the subspaces S^+ and S^- are called *half-spinor spaces*.

In (b), the type of a maximal t.s. subspace is that described in Section 7.4. The maximal t.s. subspaces containing W form a dual polar space of rank 2, which in this case is simply a complete bipartite graph, the parts of the bipartition being the two types of maximal subspace. Any two subspaces of the same type have intersection with even codimension at most 2, and hence intersect precisely in W.

The dual polar space associated with the split quadratic form has two points per line, and so in general is a bipartite graph. The two parts of the bipartition can be identified with the pure spinors in the two half-spinor spaces. The lines described in (b) within each half-spinor space form a geometry, a so-called *half-spinor geometry*: two pure spinors are collinear in this geometry if and only if they lie at distance 2 in the dual polar space. In general, distances in the half-spinor geometry are those in the dual polar space, halved!

Proposition 10.14 *If p is a point and L a line in a half-spinor geometry, then either there is a unique point of L nearest p, or all points of L are equidistant from p.*

Proof Recall that the line L of the half-spinor geometry is "half" of a complete bipartite graph Q, which is a quad in the dual polar space. If the gateway to Q is on L, it is the point of L nearest to p; if it is on the other side, then all points of L are equidistant from p. ■

The cases $n = 3, 4$ give us yet another way of looking at the Klein quadric and triality.

Example $n = 3$. The half-spinor space has rank 4. The diameter of the half-spinor geometry is 1, and so it is a linear space; necessarily $\mathrm{PG}(3,F)$: that is, every spinor in the half-spinor space is pure. Points of this space correspond to one family of maximal subspaces on the Klein quadric.

Example $n = 4$. Now the half-spinor spaces have rank 8, the same as V. The half-spinor space has diameter 2, and (by Proposition 10.14) satisfies the Buekenhout–Shult axiom. But we do not need to use the full classification of polar spaces here, since the geometry is already embedded in $\mathrm{PG}(7,F)$! We conclude that each half-spinor space is isomorphic to the original hyperbolic quadric.

We conclude by embedding a couple more dual polar spaces in projective spaces.

Proposition 10.15 *Let f be a quadratic form of rank $n-1$ on a vector space of rank $2n-1$. Then the dual polar space of F is embedded as all the points and some of the lines of the half-spinor space associated with a split quadratic form of rank n.*

Proof We can regard the given space as of the form $\mathbf{v}^\perp$, where $\mathbf{v}$ is a non-singular vector in a space carrying a split quadratic form of rank n. Now each t.s. subspace of rank $n-1$ for the given form is contained in a unique t.s. space of rank n of each type for the split form; so we have an injection from the given dual polar space to a half-spinor space. The map is onto: for if U is t.s. of rank n, then $U \cap \mathbf{c}^\perp$ has rank $n-1$. A line of the dual polar space consists of all the subspaces containing a fixed t.s. subspace of rank $n-2$, and so translates into a line of the half-spinor space, as required. ■

Proposition 10.16 *Let K be a quadratic extension of F, with Galois automorphism σ. Let V be a vector space of rank $2n$ over K, carrying a non-degenerate σ-Hermitian form b of rank n. Then the dual polar space associated with b is embeddable in a half-spinor geometry over F.*

Proof Let $H(\mathbf{v}) = b(\mathbf{v},\mathbf{v})$. Then $H(v) \in F$ for all $\mathbf{v} \in V$; and H is a quadratic form on the space V_F obtained by restricting scalars to F. (Note that V_F has rank $4n$ over F.) Now any maximal t.i. subspace for b is a maximal t.s. subspace for H of rank $2n$; so H is a split form, and we have an injection from points of the dual unitary space to pure spinors. Moreover, the intersection of any two of these maximal t.s. subspaces has even F-codimension in each; so they all have the same type, and our map goes to points of a half-spinor geometry.

A line of the dual polar space is defined by a t.i. subspace of rank $n-1$ (over K), which is t.s. of rank $2n-2$ over F; so it maps to a line of the half-spinor geometry, as required. ■

In the case $n = 3$, we have the duality between the unitary and non-split orthogonal spaces discussed in Section 8.3.

Exercise

1. (a) Prove that the set of endomorphisms of V with range contained in a subspace U is a left ideal.

(b) Prove that, if T has range U, then any endomorphism whose range is contained in U is a left multiple of T.

(c) Deduce that every left ideal of the endomorphism ring of V is of the form described in (a).

Bibliography

[A] V. I. Arnol'd, *Huygens & Barrow, Newton & Hooke*, Birkhäuser, Basel, 1990.

[B] E. Artin, *Geometric Algebra*, Interscience, New York, 1957.

[C] K. S. Brown, *Buildings*, Springer, New York, 1989.

[D] R. H. Bruck, *A Survey of Binary Systems*, Springer, Berlin, 1958.

[E] F. Buekenhout (ed.), *Handbook of Incidence Geometry*, Elsevier, Amsterdam, 1995.

[F] P. J. Cameron and J. H. van Lint, *Designs, Graphs, Codes and their Links*, London Math. Soc. Student Texts **22**, Cambridge Univ. Press, Cambridge, 1991.

[G] R. W. Carter, *Simple Groups of Lie Type*, Wiley, London, 1972.

[H] C. Chevalley, *The Algebraic Theory of Spinors and Clifford Algebras* (Collected Works Vol. 2), Springer, Berlin, 1997.

[I] P. M. Cohn, *Algebra*, Volume 1, Wiley, London, 1974.

[J] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *An ATLAS of Finite Groups*, Oxford University Press, Oxford, 1985.

[K] L. E. Dickson, *Linear Groups, with an Exposition of the Galois Field theory*, Teubner, Leipzig, 1901 (reprint Dover Publ., New York, 1958).

[L] J. Dieudonné, *La Géométrie des Groupes Classiques*, Springer, Berlin, 1955.

[M] P. B. Kleidman and M. W. Liebeck, *The Subgroup Structure of the Finite Classical Groups*, London Math. Soc. Lecture Notes **129**, Cambridge Univ. Press, Cambridge, 1990.

[N] H. Lüneburg, *Transitive Erweiterungen endlicher Permutationsgruppen*, Lecture Notes in Math. **84**, Springer, Berlin, 1969.

[O] G. Pickert, *Projektive Ebenen*, Springer-Verlag, Berlin, 1955.

[P] M. A. Ronan, *Lectures on Buildings*, Academic Press, Boston, 1989.

[Q] SOCRATES lecture notes, available from
`http://dwispc8.vub.ac.be/Potenza/lectnotes.html`
and
`http://cage.rug.ac.be/~fdc/intensivecourse2/final.html`

[R] D. E. Taylor, *The Geometry of the Classical Groups*, Heldermann Verlag, Berlin, 1992.

[S] J. Tits, *Buildings of Spherical Type and Finite BN-Pairs*, Lecture Notes in Math. **386**, Springer, Berlin, 1974.

[T] O. Veblen and J. W. Young (1916), *Projective Geometry* (2 vols.), Ginn & Co., Boston, 1916.

---

[1] A. Barlotti, Sul gruppo delle proiectività di una retta in se nei piani liberi e nei piani aperti, *Rendic. Sem. Math. Padova* **34** (1964), 135–139.

[2] D. R. Breach and A. P. Street, Partitioning sets of quadruples into designs, II, *J. Comb. Math. Comb. Comput.* **3** (1988), 41-48.

[3] M. Brown, (Hyper)ovals and ovoids in projective spaces, available from
`http://cage.rug.ac.be/~fdc/intensivecourse2/brown_2.pdf`

[4] R. H. Bruck and H. J. Ryser, The nonexistence of certain finite projective planes, *Canad. J. Math.* **1** (1949), 88–93.

[5] F. Buekenhout, Plans projectifs à ovoïdes pascaliens, *Arch. Math.* **17** (1966), 89–93.

[6] F. Buekenhout, Une caractérisation des espaces affines basée sur la notion de droite, *Math. Z.* **111** (1969), 367–371.

[7] F. Buekenhout, Diagrams for geometries and groups, *J. Combinatorial Theory* (A) **27** (1979), 121–151.

[8] F. Buekenhout and E. E. Shult (1974), On the foundations of polar geometry, *Geometriae Dedicata* **3** (1974), 155–170.

[9] P. J. Cameron, Dual polar spaces, *Geometriae dedicata* **12** (1982), 75–85.

[10] P. J. Cameron, *Classical Groups*, available from
`http://www.maths.qmw.ac.uk/~pjc/class_gps/`

[11] P. J. Cameron and J. I. Hall, Some groups generated by transvection subgroups, *J. Algebra* **140** (1991), 184–209.

[12] P. J. Cameron and W. M. Kantor, 2-transitive and antiflag transitive collineation groups of finite projective spaces, *J. Algebra* **60** (1979), 384–422.

[13] P. J. Cameron and C. E. Praeger, Partitioning into Steiner systems, in *Combinatorics '88*, Mediterranean Press, Roma, 1991.

[14] W. Cherowitzo, Hyperoval page, available from

`http://www-math.cudenver.edu/~wcherowi/research/hyperoval/hypero.html`

[15] P. M. Cohn, The embedding of firs in skew fields, *Proc. London Math. Soc.* (3) **23** (1971), 193–213.

[16] J. Doyen and X. Hubaut, Finite regular locally projective spaces, *Math. Z.* **119** (1971), 83–88.

[17] W. Feit and G. Higman, On the non-existence of certain generalised polygons, *J. Algebra* **1** (1964), 434–446.

[18] M. Hall Jr., Automorphisms of Steiner triple systems, *IBM J. Res. Develop.* **4** (1960), 460–471.

[19] M. Hall Jr., Group theory and block designs, pp. 115-144 in *Proc. Internat. Conf. theory of Groups* (ed. L. G. Kovács and B. H. Neumann), Gordon & Breach, New York, 1967.

[20] G. Higman, B. H. Neumann and H. Neumann, Embedding theorems for groups, *J. London Math. Soc.* (1) **26** (1949), 247–254.

[21] C. W. H. Lam, S. Swiercz and L. Thiel, The nonexistence of finite projective planes of order 10, *Canad. J. Math.* **41** (1989), 1117–1123.

[22] J. E. McLaughlin, Some groups generated by transvections, *Arch. Math. (Basel)* **18** (1967), 362–368.

[23] J. E. McLaughlin, Some subgroups of $SL_n(F_2)$, *Illinois J. Math.* **13** (1969), 108–115.

[24] R. Scharlau, Buildings, pp. 477–645 in *Handbook of Incidence Geometry* (F. Buekenhout, ed.), Elsevier, Amsterdam, 1995.

[25] A. Schleiermacher, Bemerkungen zum Fundamentalsatz der projectivern Geometrie, *Math. Z.* **99** (1967), 299–304.

[26] B. Segre, Sulle ovali nei piani lineari finiti, *Atti Accad. Naz. Lincei Rendic.* **17** (1954), 141–142.

[27] J. J. Seidel, On two-graphs, and Shult's characterization of symplectic and orthogonal geometries over GF(2), *T.H. report 73-WSK-02*, Techn. Univ., Eindhoven, 1973.

[28] E. E. Shult, Characterizations of certain classes of graphs, *J. Combinatorial Theory* (B) **13** (1972), 142–167.

[29] E. E. Shult, Characterizations of the Lie incidence geometries, pp. 157–186 in *Surveys in Combinatorics* (ed. E. K. Lloyd), London Math. Soc. Lecture Notes **82**, Cambridge Univ. Press, Cambridge, 1983.

[30] E. E. Shult and A. Yanushka, Near n-gons and line systems, *Geometriae Dedicata*, **9** (1980), 1–72.

[31] A. P. Sprague, Pasch's axiom and projective spaces, *Discrete Math.* **33** (1981), 79–87.

[32] J. A. Thas, P. J. Cameron and A. Blokhuis, On a generalization of a theorem of B. Segre, *Geometriae Dedicata* **43** (1992), 299–305.

[33] J. Tits, Sur la trialité et certains groupes qui s'en déduisent, *Publ. Math. I.H.E.S.* **2** (1959), 14–60.

[34] J. Tits, Ovoïdes et groupes de Suzuki, *Arch. Math.* **13** (1962), 187–198.

[35] J. Tits, Nonexistence de certains polygons généralisés, I, *Invent. Math.* **36** (1976), 275–284.

[36] J. Tits, Nonexistence de certains polygons généralisés, II, *Invent. Math.* **51** (1979), 267–269.

[37] J. A. Todd, As it might have been, *Bull. London Math. Soc.* **2** (1970), 1–4.

[38] O. Veblen and J. H. M. Wedderburn, Non-Desarguesian and non-Pascalian geometries, *Trans. Amer. Math. Soc.* **8** (1907), 379–388.

[39] R. Weiss, The nonexistence of certain Moufang polygons, *Invent. Math.* **51** (1979), 261–266.

Index