Notes on Classical Groups

Peter J. Cameron
School of Mathematical Sciences
Queen Mary and Westfield College
London E1 4NS
U.K.
p.j.cameron@qmw.ac.uk

These notes are the content of an M.Sc. course I gave at Queen Mary and Westfield College, London, in January–March 2000.

I am grateful to the students on the course for their comments; to Keldon Drudge, for standing in for me; and to Simeon Ball, for helpful discussions.

Contents:

1 Fields and vector spaces

In this section we revise some algebraic preliminaries and establish notation.

1.1 Division rings and fields

A *division ring*, or *skew field*, is a structure F with two binary operations called *addition* and *multiplication*, satisfying the following conditions:

(a) $(F,+)$ is an abelian group, with identity 0, called the *additive group* of F;

(b) $(F \setminus 0, \cdot)$ is a group, called the *multiplicative group* of F;

(c) left or right multiplication by any fixed element of F is an endomorphism of the additive group of F.

Note that condition (c) expresses the two distributive laws. Note that we must assume both, since one does not follow from the other.

The identity element of the multiplicative group is called 1.

A *field* is a division ring whose multiplication is commutative (that is, whose multiplicative group is abelian).

Exercise 1.1 Prove that the commutativity of addition follows from the other axioms for a division ring (that is, we need only assume that $(F,+)$ is a group in (a)).

Exercise 1.2 A *real quaternion* has the form $a+b\mathrm{i}+c\mathrm{j}+d\mathrm{k}$, where $a,b,c,d \in \mathbb{R}$. Addition and multiplication are given by "the usual rules", together with the following rules for multiplication of the elements $1,\mathrm{i},\mathrm{j},\mathrm{k}$:

$\cdot$	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-$j
j	j	$-$k	-1	i
k	k	j	$-$i	-1

Prove that the set $\mathbb{H}$ of real quaternions is a division ring. (*Hint*: If $q = a+b\mathrm{i}+c\mathrm{j}+d\mathrm{k}$, let $q^* = a-b\mathrm{i}-c\mathrm{j}-d\mathrm{k}$; prove that $qq^* = a^2+b^2+c^2+d^2$.)

Multiplication by zero induces the zero endomorphism of $(F,+)$. Multiplication by any non-zero element induces an automorphism (whose inverse is multiplication by the inverse element). In particular, we see that the automorphism group of $(F,+)$ acts transitively on its non-zero elements. So all non-zero elements have the same order, which is either infinite or a prime p. In the first case, we say that the *characteristic* of F is zero; in the second case, it has *characteristic* p.

The structure of the multiplicative group is not so straightforward. However, the possible finite subgroups can be determined. If F is a field, then any finite subgroup of the multiplicative group is cyclic. To prove this we require *Vandermonde's Theorem*:

Theorem 1.1 *A polynomial equation of degree n over a field has at most n roots.*

Exercise 1.3 Prove Vandermonde's Theorem. (*Hint*: If $f(a) = 0$, then $f(x) = (x-a)g(x)$.)

Theorem 1.2 *A finite subgroup of the multiplicative group of a field is cyclic.*

Proof An element ω of a field F is an *nth root of unity* if $\omega^n = 1$; it is a *primitive* nth root of unity if also $\omega^m \neq 1$ for $0 < m < n$.

Let G be a subgroup of order n in the multiplicative group of the field F. By Lagrange's Theorem, every element of G is an nth root of unity. If G contains a primitive nth root of unity, then it is cyclic, and the number of primitive nth roots is $\phi(n)$, where ϕ is Euler's function. If not, then of course the number of primitive nth roots is zero. The same considerations apply of course to any divisor of n. So, if $\psi(m)$ denotes the number of primitive mth roots of unity in G, then

(a) for each divisor m of n, either $\psi(m) = \phi(m)$ or $\psi(m) = 0$.

Now every element of G has some finite order dividing n; so

(b) $\sum_{m|n} \psi(m) = n$.

Finally, a familiar property of Euler's function yields:

(c) $\sum_{m|n} \phi(m) = n$.

From (a), (b) and (c) we conclude that $\psi(m) = \phi(m)$ for all divisors m of n. In particular, $\psi(n) = \phi(n) \neq 0$, and G is cyclic. ∎

For division rings, the position is not so simple, since Vandermonde's Theorem fails.

Exercise 1.4 Find all solutions of the equation $x^2+1=0$ in $\mathbb{H}$.

However, the possibilities can be determined. Let G be a finite subgroup of the multiplicative group of the division ring F. We claim that there is an abelian group A such that G is a group of automorphisms of A acting semiregularly on the non-zero elements. Let B be the subgroup of $(F,+)$ generated by G. Then B is a finitely generated abelian group admitting G acting semiregularly. If F has non-zero characteristic, then B is elementary abelian; take $A=B$. Otherwise, choose a prime p such that, for all $x,g\in G$, the element $(xg-x)p^{-1}$ is not in B, and set $A=B/pB$.

The structure of semiregular automorphism groups of finite groups (a.k.a. *Frobenius complements*) was determined by Zassenhaus. See Passman, *Permutation Groups*, Benjamin, New York, 1968, for a detailed account. In particular, either G is metacyclic, or it has a normal subgroup isomorphic to $\mathrm{SL}(2,3)$ or $\mathrm{SL}(2,5)$. (These are finite groups G having a unique subgroup Z of order 2, such that G/Z is isomorphic to the alternating group A_4 or A_5 respectively. There is a unique such group in each case.)

Exercise 1.5 Identify the division ring $\mathbb{H}$ of real quaternions with the real vector space $\mathbb{R}^4$ with basis $\{1,\mathrm{i},\mathrm{j},\mathrm{k}\}$. Let U denote the multiplicative group of *unit quaternions*, those elements $a+b\mathrm{i}+c\mathrm{j}+d\mathrm{k}$ satisfying $a^2+b^2+c^2+d^2=1$. Show that conjugation by a unit quaternion is an orthogonal transformation of $\mathbb{R}^4$, fixing the 1-dimensional space spanned by 1 and inducing an orthogonal transformation on the 3-dimensional subspace spanned by $\mathrm{i},\mathrm{j},\mathrm{k}$.

Prove that the map from U to the 3-dimensional orthogonal group has kernel ± 1 and image the group of rotations of 3-space (orthogonal transformations with determinant 1).

Hence show that the groups $\mathrm{SL}(2,3)$ and $\mathrm{SL}(2,5)$ are finite subgroups of the multiplicative group of $\mathbb{H}$.

Remark: This construction explains why the groups $\mathrm{SL}(2,3)$ and $\mathrm{SL}(2,5)$ are sometimes called the *binary tetrahedral* and *binary icosahedral* groups. Construct also a *binary octahedral* group of order 48, and show that it is not isomorphic to $\mathrm{GL}(2,3)$ (the group of 2×2 invertible matrices over the integers mod 3), even though both groups have normal subgroups of order 2 whose factor groups are isomorphic to the symmetric group S_4.

1.2 Finite fields

The basic facts about finite fields are summarised in the following two theorems, due to Wedderburn and Galois respectively.

Theorem 1.3 *Every finite division ring is commutative.*

Theorem 1.4 *The number of elements in a finite field is a prime power. Conversely, if q is a prime power, then there is a unique field with q elements, up to isomorphism.*

The unique finite field with a given prime power order q is called the *Galois field* of order q, and denoted by $\mathrm{GF}(q)$ (or sometimes $\mathbb{F}_q$). If q is prime, then $\mathrm{GF}(q)$ is isomorphic to $\mathbb{Z}/q\mathbb{Z}$, the integers mod q.

We now summarise some results about $\mathrm{GF}(q)$.

Theorem 1.5 *Let $q = p^a$, where p is prime and a is a positive integer. Let $F =$* $\mathrm{GF}(q)$.

(a) *F has characteristic p, and its additive group is an elementary abelian p-group.*

(b) *The multiplicative group of F is cyclic, generated by a primitive $(p^a - 1)$th root of unity (called a* primitive element *of F).*

(c) *The automorphism group of F is cyclic of order a, generated by the* Frobenius automorphism $x \mapsto x^p$.

(d) *For every divisor b of a, there is a unique subfield of F of order p^b, consisting of all solutions of $x^{p^b} = x$; and these are all the subfields of F.*

Proof Part (a) is obvious since the additive group contains an element of order p, and part (b) follows from Theorem 1.2. Parts (c) and (d) are most easily proved using Galois theory. Let E denote the subfield $\mathbb{Z}/p\mathbb{Z}$ of F. Then the degree of F over E is a. The Frobenius map $\sigma : x \mapsto x^p$ is an E-automorphism of F, and has order a; so F is a Galois extension of E, and σ generates the Galois group. Now subfields of F necessarily contain E; by the Fundamental Theorem of Galois Theory, they are the fixed fields of subgroups of the Galois group $\langle\sigma\rangle$. ■

For explicit calculation in $F = \mathrm{GF}(p^a)$, it is most convenient to represent it as $E[x]/(f)$, where $E = \mathbb{Z}/p\mathbb{Z}$, $E[x]$ is the polynomial ring over E, and f is the (irreducible) minimum polynomial of a primitive element of F. If α denotes the coset $(f)+x$, then α is a root of f, and hence a primitive element.

Now every element of F can be written uniquely in the form

$$c_0 + c_1\alpha + \cdots + c_{a-1}\alpha^{a-1},$$

where $c_0, c_1, \ldots, c_{a-1} \in E$; addition is straightforward in this representation. Also, every non-zero element of F can be written uniquely in the form α^m, where $0 \le m < p^a - 1$, since α is primitive; multiplication is straightforward in this representation. Using the fact that $f(\alpha) = 0$, it is possible to construct a table matching up the two representations.

Example The polynomial $x^3 + x + 1$ is irreducible over $E = \mathbb{Z}/2\mathbb{Z}$. So the field $F = E(\alpha)$ has eight elements, where α satisfies $\alpha^3 + \alpha + 1 = 0$ over E. We have $\alpha^7 = 1$, and the table of logarithms is as follows:

α^0	1
α^1	α
α^2	α^2
α^3	$\alpha + 1$
α^4	$\alpha^2 + \alpha$
α^5	$\alpha^2 + \alpha + 1$
α^6	$\alpha^2 + 1$

Hence

$$(\alpha^2 + \alpha + 1)(\alpha^2 + 1) = \alpha^5 \cdot \alpha^6 = \alpha^4 = \alpha^2 + \alpha.$$

Exercise 1.6 Show that there are three irreducible polynomials of degree 4 over the field $\mathbb{Z}/2\mathbb{Z}$, of which two are primitive. Hence construct GF(16) by the method outlined above.

Exercise 1.7 Show that an irreducible polynomial of degree m over $\mathrm{GF}(q)$ has a root in $\mathrm{GF}(q^n)$ if and only if m divides n.

Hence show that the number a_m of irreducible polynomials of degree m over $\mathrm{GF}(q)$ satisfies

$$\sum_{m|n} m a_m = q^n.$$

Exercise 1.8 Show that, if q is even, then every element of $\mathrm{GF}(q)$ is a square; while, if q is odd, then half of the non-zero elements of $\mathrm{GF}(q)$ are squares and half are non-squares.

If q is odd, show that -1 is a square in $\mathrm{GF}(q)$ if and only if $q \equiv 1 \pmod 4$.

1.3 Vector spaces

A *left vector space* over a division ring F is a unital left F-module. That is, it is an abelian group V, with a anti-homomorphism from F to $\mathrm{End}(V)$ mapping 1 to the identity endomorphism of V.

Writing scalars on the left, we have $(cd)v = c(dv)$ for all $c,d \in F$ and $v \in V$: that is, scalar multiplication by cd is the same as multiplication by d followed by multiplication by c, not vice versa. (The opposite convention would make V a right (rather than left) vector space; scalars would more naturally be written on the right.) The unital condition simply means that $1v = v$ for all $v \in V$.

Note that F is a vector space over itself, using field multiplication for the scalar multiplication.

If F is a division ring, the *opposite* division ring F° has the same underlying set as F and the same addition, with multiplication given by

$$a \circ b = ba.$$

Now a right vector space over F can be regarded as a left vector space over F°.

A *linear transformation* $T : V \to W$ between two left F-vector spaces V and W is a vector space homomorphism; that is, a homomorphism of abelian groups which commutes with scalar multiplication. We write linear transformations on the right, so that we have

$$(cv)T = c(vT)$$

for all $c \in F$, $v \in V$. We add linear transformations, or multiply them by scalars, pointwise (as functions), and multiply then by function composition; the results are again linear transformations.

If a linear transformation T is one-to-one and onto, then the inverse map is also a linear transformation; we say that T is *invertible* if this occurs.

Now $\mathrm{Hom}(V,W)$ denotes the set of all linear transformations from V to W. The *dual space* of F is $F^* = \mathrm{Hom}(V,F)$.

Exercise 1.9 Show that V^* is a right vector space over F.

A vector space is *finite-dimensional* if it is finitely generated as F-module. A *basis* is a minimal generating set. Any two bases have the same number of elements; this number is usually called the dimension of the vector space, but in order to avoid confusion with a slightly different geometric notion of dimension, I will call it the *rank* of the vector space. The rank of V is denoted by $\mathrm{rk}(V)$.

Every vector can be expressed uniquely as a linear combination of the vectors in a basis. In particular, a linear combination of basis vectors is zero if and only if all the coefficients are zero. Thus, a vector space of rank n over F is isomorphic to F^n (with coordinatewise addition and scalar multiplication).

I will assume familiarity with standard results of linear algebra about ranks of sums and intersections of subspaces, about ranks of images and kernels of linear transformations, and about the representation of linear transformations by matrices with respect to given bases.

As well as linear transformations, we require the concept of a *semilinear transformation* between F-vector spaces V and W. This can be defined in two ways. It is a map T from V to W satisfying

(a) $(v_1 + v_2)T = v_1T + v_2T$ for all $v_1, v_2 \in V$;

(b) $(cv)T = c^\sigma vT$ for all $c \in F$, $v \in V$, where σ is an automorphism of F called the *associated automorphism* of T.

Note that, if T is not identically zero, the associated automorphism is uniquely determined by T.

The second definition is as follows. Given an automorphism σ of F, we extend the action of σ to F^n coordinatewise:

$$(c_1, \ldots, c_n)^\sigma = (c_1^\sigma, \ldots, c_n^\sigma).$$

Hence we have an action of σ on any F-vector space with a given basis. Now a σ-*semilinear transformation* from V to W is the composition of a linear transformation from V to W with the action of σ on W (with respect to some basis).

The fact that the two definitions agree follows from the observations

- the action of σ on F^n is semilinear in the first sense;
- the composition of semilinear transformations is semilinear (and the associated automorphism is the composition of the associated automorphisms of the factors).

This immediately shows that a semilinear map in the second sense is semilinear in the first. Conversely, if T is semilinear with associated automorphism σ, then the composition of T with σ^{-1} is linear, so T is σ-semilinear.

Exercise 1.10 Prove the above assertions.

If a semilinear transformation T is one-to-one and onto, then the inverse map is also a semilinear transformation; we say that T is *invertible* if this occurs.

Almost exclusively, I will consider only finite-dimensional vector spaces. To complete the picture, here is the situation in general. In ZFC (Zermelo–Fraenkel set theory with the Axiom of Choice), every vector space has a basis (a set of vectors with the property that every vector has a unique expression as a linear combination of a *finite* set of basis vectors with non-zero coefficients), and any two bases have the same cardinal number of elements. However, without the Axiom of Choice, there may exist a vector space which has no basis.

Note also that there exist division rings F with bimodules V such that V has different ranks when regarded as a left or a right vector space.

1.4 Projective spaces

It is not easy to give a concise definition of a projective space, since projective geometry means several different things: a geometry with points, lines, planes, and so on; a topological manifold with a strange kind of torsion; a lattice with meet, join, and order; an abstract incidence structure; a tool for computer graphics.

Let V be a vector space of rank $n+1$ over a field F. The "objects" of the n-dimensional projective space are the subspaces of V, apart from V itself and the zero subspace $\{0\}$. Each object is assigned a dimension which is one less than its rank, and we use geometric terminology, so that *points*, *lines* and *planes* are the objects of dimension 0, 1 and 2 (that is, rank 1, 2, 3 respectively). A *hyperplane* is an object having codimension 1 (that is, dimension $n-1$, or rank n). Two objects are *incident* if one contains the other. So two objects of the same dimension are incident if and only if they are equal.

The n-dimensional projective space is denoted by $\mathrm{PG}(n,F)$. If F is the Galois field $\mathrm{GF}(q)$, we abbreviate $\mathrm{PG}(n,\mathrm{GF}(q))$ to $\mathrm{PG}(n,q)$. A similar convention will be used for other geometries and groups over finite fields.

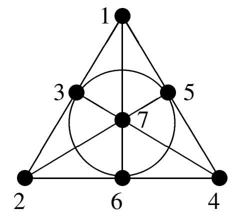
A 0-dimensional projective space has no internal structure at all, like an idealised point. A 1-dimensional projective space is just a set of points, one more than the number of elements of F, with (at the moment) no further structure. (If

$\{e_1, e_2\}$ is a basis for V, then the points are spanned by the vectors $\lambda e_1 + e_2$ (for $\lambda \in F$) and e_1.)

For $n > 1$, $\mathrm{PG}(n,F)$ contains objects of different dimensions, and the relation of incidence gives it a non-trivial structure.

Instead of our "incidence structure" model, we can represent a projective space as a collection of subsets of a set. Let S be the set of points of $\mathrm{PG}(n,F)$. The *point shadow* of an object U is the set of points incident with U. Now the point shadow of a point P is simply $\{P\}$. Moreover, two objects are incident if and only if the point shadow of one contains that of the other.

The diagram below shows $\mathrm{PG}(2,2)$. It has seven points, labelled 1, 2, 3, 4, 5, 6, 7; the line shadows are 123, 145, 167, 246, 257, 347 356 (where, for example, 123 is an abbreviation for $\{1,2,3\}$).



The correspondence between points and spanning vectors of the rank-1 subspaces can be taken as follows:

1	2	3	4	5	6	7
$(0,0,1)$	$(0,1,0)$	$(0,1,1)$	$(1,0,0)$	$(1,0,1)$	$(1,1,0)$	$(1,1,1)$

The following geometric properties of projective spaces are easily verified from the rank formulae of linear algebra:

(a) Any two distinct points are incident with a unique line.

(b) Two distinct lines contained in a plane are incident with a unique point.

(c) Any three distinct points, or any two distinct collinear lines, are incident with a unique plane.

(d) A line not incident with a given hyperplane meets it in a unique point.

(e) If two distinct points are both incident with some object of the projective space, then the unique line incident with them is also incident with that object.

Exercise 1.11 Prove the above assertions.

It is usual to be less formal with the language of incidence, and say "the point P lies on the line L", or "the line L passes through the point P" rather than "the point P and the line L are incident". Similar geometric language will be used without further comment.

An *isomorphism* from a projective space Π_1 to a projective space Π_2 is a map from the objects of Π_1 to the objects of Π_2 which preserves the dimensions of objects and also preserves the relation of incidence between objects. A *collineation* of a projective space Π is an isomorphism from Π to Π.

The important theorem which connects this topic with that of the previous section is the *Fundamental Theorem of Projective Geometry*:

Theorem 1.6 *Any isomorphism of projective spaces of dimension at least two is induced by an invertible semilinear transformation of the underlying vector spaces. In particular, the collineations of* $\mathrm{PG}(n,F)$ *for* $n \geq 2$ *are induced by invertible semilinear transformations of the rank-*$(n+1)$ *vector space over* F.

This theorem will not be proved here, but I make a few comments about the proof. Consider first the case $n = 2$. One shows that the field F can be recovered from the projective plane (that is, the addition and multiplication in F can be defined by geometric constructions involving points and lines). The construction is based on choosing four points of which no three are collinear. Hence any collineation fixing these four points is induced by a field automorphism. Since the group of invertible linear transformations acts transitively on quadruples of points with this property, it follows that any collineation is induced by the composition of a linear transformation and a field automorphism, that is, a semilinear transformation.

For higher-dimensional spaces, we show that the coordinatisations of the planes fit together in a consistent way to coordinatise the whole space.

In the next chapter we study properties of the collineation group of projective spaces. Since we are concerned primarily with groups of matrices, I will normally speak of $\mathrm{PG}(n-1,F)$ as the projective space based on a vector space of rank n, rather than $\mathrm{PG}(n,F)$ based on a vector space of rank $n+1$.

Next we give some numerical information about finite projective spaces.

Theorem 1.7 *(a) The number of points in the projective space* $\mathrm{PG}(n-1,q)$ *is* $(q^n-1)/(q-1)$.

(b) *More generally, the number of* $(m-1)$*-dimensional subspaces of* $\mathrm{PG}(n-1,q)$ *is*

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

(c) *The number of* $(m-1)$*-dimensional subspaces of* $\mathrm{PG}(n-1,q)$ *containing a given* $(l-1)$*-dimensional subspace is equal to the number of* $(m-l-1)$*-dimensional subspaces of* $\mathrm{PG}(n-l-1,q)$.

Proof (a) The projective space is based on a vector space of rank n, which contains q^n vectors. One of these is the zero vector, and the remaining q^n-1 each span a subspace of rank 1. Each rank 1 subspace contains $q-1$ non-zero vectors, each of which spans it.

(b) Count the number of linearly independent m-tuples of vectors. The jth vector must lie outside the rank $(j-1)$ subspace spanned by the preceding vectors, so there are q^n-q^{j-1} choices for it. So the number of such m-tuples is the numerator of the fraction. By the same argument (replacing n by m), the number of linearly independent m-tuples which span a given rank m subspace is the denominator of the fraction.

(c) If U is a rank l subspace of the rank m vector space V, then the Second Isomorphism Theorem shows that there is a bijection between rank m subspaces of V containing U, and rank $(m-l)$ subspaces of the rank $(n-l)$ vector space V/U. $\blacksquare$

The number given by the fraction in part (b) of the theorem is called a *Gaussian coefficient*, written $\begin{bmatrix} n \\ m \end{bmatrix}_q$. Gaussian coefficients have properties resembling those of binomial coefficients, to which they tend as $q \to 1$.

Exercise 1.12 (a) Prove that

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

(b) Prove that for $n \geq 1$,

$$\prod_{i=0}^{n-1}(1+q^i x) = \sum_{k=0}^{n} q^{k(k-1)/2}\begin{bmatrix} n \\ k \end{bmatrix}_q x^k.$$

(This result is known as the *q-binomial theorem*, since it reduces to the binomial theorem as $q \to 1$.)

If we regard a projective space $\mathrm{PG}(n-1,F)$ purely as an incidence structure, the dimensions of its objects are not uniquely determined. This is because there is an additional symmetry known as *duality*. That is, if we regard the hyperplanes as points, and define new dimensions by $\dim^*(U) = n-2-\dim(U)$, we again obtain a projective space, with the same relation of incidence. The reason that it is a projective space is as follows.

Let $V^* = \mathrm{Hom}(V,F)$ be the dual space of V, where V is the underlying vector space of $\mathrm{PG}(n-1,F)$. Recall that V^* is a right vector space over F, or equivalently a left vector space over the opposite field F°. To each subspace U of V, there is a corresponding subspace $U^\dagger$ of V^*, the *annihilator* of U, given by

$$U^\dagger = \{f \in V^* : uf = 0 \text{ for all } u \in U\}.$$

The correspondence $U \mapsto U^\dagger$ is a bijection between the subspaces of V and the subspaces of V^*; we denote the inverse map from subspaces of V^* to subspaces of V also by $\dagger$. It satisfies

(a) $(U^\dagger)^\dagger = U$;

(b) $U_1 \leq U_2$ if and only if $U_1^\dagger \geq U_2^\dagger$;

(c) $\mathrm{rk}(U^\dagger) = n - \mathrm{rk}(U)$.

Thus we have:

Theorem 1.8 *The dual of* $\mathrm{PG}(n-1,F)$ *is the projective space* $\mathrm{PG}(n-1,F^\circ)$. *In particular, if $n \geq 3$, then* $\mathrm{PG}(n-1,F)$ *is isomorphic to its dual if and only if F is isomorphic to its opposite F°.*

Proof The first assertion follows from our remarks. The second follows from the first by use of the Fundamental Theorem of Projective Geometry. ∎

Thus, $\mathrm{PG}(n-1,F)$ is self-dual if F is commutative, and for some non-commutative division rings such as $\mathbb{H}$; but there are division rings F for which $F \not\cong F^\circ$.

An isomorphism from F to its opposite is a bijection σ satisfying

$$\begin{aligned}(a+b)^\sigma &= a^\sigma + b^\sigma,\\ (ab)^\sigma &= b^\sigma a^\sigma,\end{aligned}$$

for all $a,b \in F$. Such a map is called an *anti-automorphism* of F.

Exercise 1.13 Show that $\mathbb{H} \cong \mathbb{H}^\circ$. (*Hint*: $(a+b\mathrm{i}+c\mathrm{j}+d\mathrm{k})^\sigma = a - b\mathrm{i} - c\mathrm{j} - d\mathrm{k}$.)

2 Linear and projective groups

In this section, we define and study the general and special linear groups and their projective versions. We look at the actions of the projective groups on the points of the projective space, and discuss transitivity properties, generation, and simplicity of these groups.

2.1 The general linear groups

Let F be a division ring. As we saw, a vector space of rank n over F can be identified with the standard space F^n (with scalars on the left) by choosing a basis. Any invertible linear transformation of V is then represented by an invertible $n \times n$ matrix, acting on F^n by right multiplication.

We let $\mathrm{GL}(n,F)$ denote the group of all invertible $n \times n$ matrices over F, with the operation of matrix multiplication.

The group $\mathrm{GL}(n,F)$ acts on the projective space $\mathrm{PG}(n-1,F)$, since an invertible linear transformation maps a subspace to another subspace of the same dimension.

Proposition 2.1 *The kernel of the action of* $\mathrm{GL}(n,F)$ *on the set of points of* $\mathrm{PG}(n-1,F)$ *is the subgroup*

$$\{cI : c \in Z(F), c \neq 0\}$$

of central scalar matrices in F, where $Z(F)$ *denotes the centre of* F.

Proof Let $A = (a_{ij})$ be an invertible matrix which fixes every rank 1 subspace of F^n. Thus, A maps each non-zero vector $(x_1,\ldots,x_n)$ to a scalar multiple $(cx_1,\ldots,cx_n)$ of itself.

Let e_i be the ith basis vector, with 1 in position i and 0 elsewhere. Then $e_iA = c_ie_i$, so the ith row of A is c_ie_i. This shows that A is a diagonal matrix.

Now for $i \neq j$, we have

$$c_ie_i + c_je_j = (e_i + e_j)A = d(e_i + e_j)$$

for some d. So $c_i = c_j$. Thus, A is a diagonal matrix cI.

Finally, let $a \in F$, $a \neq 0$. Then

$$c(ae_1) = (ae_1)A = a(e_1A) = ace_1,$$

so $ac = ca$. Thus, $c \in Z(F)$. ∎

Let Z be the kernel of this action. We define the *projective general linear group* $\mathrm{PGL}(n,F)$ to be the group induced on the points of the projective space $\mathrm{PG}(n-1,F)$ by $\mathrm{GL}(n,F)$. Thus,

$$\mathrm{PGL}(n,F) \cong \mathrm{GL}(n,F)/Z.$$

In the case where F is the finite field $\mathrm{GF}(q)$, we write $\mathrm{GL}(n,q)$ and $\mathrm{PGL}(n,q)$ in place of $\mathrm{GL}(n,F)$ and $\mathrm{PGL}(n,F)$ (with similar conventions for the groups we meet later). Now we can compute the orders of these groups:

Theorem 2.2 *(a)* $|\mathrm{GL}(n,q)| = (q^n-1)(q^n-q)\cdots(q^n-q^{n-1})$*;*

(b) $|\mathrm{PGL}(n,q)| = |\mathrm{GL}(n,q)|/(q-1)$.

Proof (a) The rows of an invertible matrix over a field are linearly independent, that is, for $i = 1,\ldots,n$, the ith row lies outside the subspace of rank $i-1$ generated by the preceding rows. Now the number of vectors in a subspace of rank $i-1$ over $\mathrm{GF}(q)$ is q^{i-1}, so the number of choices for the ith row is $q^n - q^{i-1}$. Multiplying these numbers for $i = 1,\ldots,n$ gives the result.

(b) $\mathrm{PGL}(n,q)$ is the image of $\mathrm{GL}(n,q)$ under a homomorphism whose kernel consists of non-zero scalar matrices and so has order $q-1$. ■

If the field F is commutative, then the determinant function is defined on $n\times n$ matrices over F and is a multiplicative map to F:

$$\det(AB) = \det(A)\det(B).$$

Also, $\det(A) \neq 0$ if and only if A is invertible. So det is a homomorphism from $\mathrm{GL}(n,F)$ to F^*, the multiplicative group of F (also known as $\mathrm{GL}(1,F)$). This homomorphism is onto, since the matrix with c in the top left corner, 1 in the other diagonal positions, and 0 elsewhere has determinant c.

The kernel of this homomorphism is the *special linear group* $\mathrm{SL}(n,F)$, a normal subgroup of $\mathrm{GL}(n,F)$ with factor group isomorphic to F^*.

We define the *projective special linear group* $\mathrm{PSL}(n,F)$ to be the image of $\mathrm{SL}(n,F)$ under the homomorphism from $\mathrm{GL}(n,F)$ to $\mathrm{PGL}(n,F)$, that is, the group induced on the projective space by $\mathrm{SL}(n,F)$. Thus,

$$\mathrm{PSL}(n,F) = \mathrm{SL}(n,F)/(SL(n,F)\cap Z).$$

The kernel of this homomorphism consists of the scalar matrices cI which have determinant 1, that is, those cI for which $c^n = 1$. This is a finite cyclic group whose order divides n.

Again, for finite fields, we can calculate the orders:

Theorem 2.3 *(a)* $|\mathrm{SL}(n,q)| = |\mathrm{GL}(n,q)|/(q-1)$*;*

(b) $|\mathrm{PSL}(n,q)| = |\mathrm{SL}(n,q)|/(n,q-1)$*, where* $(n,q-1)$ *is the greatest common divisor of n and* $q-1$.

Proof (a) $\mathrm{SL}(n,q)$ is the kernel of the determinant homomorphism on $\mathrm{GL}(n,q)$ whose image F^* has order $q-1$.

(b) From the remark before the theorem, we see that $\mathrm{PSL}(n,q)$ is the image of $\mathrm{SL}(n,q)$ under a homomorphism whose kernel is the group of nth roots of unity in $\mathrm{GF}(q)$. Since the multiplicative group of this field is cyclic of order $q-1$, the nth roots form a subgroup of order $(n,q-1)$. ■

A group G acts *sharply transitively* on a set Ω if its action is regular, that is, it is transitive and the stabiliser of a point is the identity.

Theorem 2.4 *Let F be a division ring. Then the group* $\mathrm{PGL}(n,F)$ *acts transitively on the set of all* $(n+1)$*-tuples of points of* $\mathrm{PG}(n-1,F)$ *with the property that no n points lie in a hyperplane; the stabiliser of such a tuple is isomorphic to the group of inner automorphisms of the multiplicative group of F. In particular, if F is commutative, then* $\mathrm{PGL}(n,F)$ *is sharply transitive on the set of such* $(n+1)$*-tuples.*

Proof Consider n points not lying in a hyperplane. The n vectors spanning these points form a basis, and we may assume that this is the standard basis $e_1,\ldots,e_n$ of F^n, where e_i has ith coordinate 1 and all others zero. The proof of Proposition 2.1 shows that G acts transitively on the set of such n-tuples, and the stabiliser of the n points is the group of diagonal matrices. Now a vector v not lying in the hyperplane spanned by any $n-1$ of the basis vectors must have all its coordinates non-zero, and conversely. Moreover, the group of diagonal matrices acts transitively on the set of such vectors. This proves that $\mathrm{PG}(n,F)$ is transitive on the set of $(n+1)$-tuples of the given form. Without loss of generality, we may assume that $v = e_1 + \cdots + e_n = (1,1,\ldots,1)$. Then the stabiliser of the $n+1$ points consists of the group of scalar matrices, which is isomorphic to the multiplicative group F^*. We have seen that the kernel of the action on the projective space is $Z(F^*)$, so the group induced by the scalar matrices is $F^*/Z(F^*)$, which is isomorphic to the group of inner automorphisms of F^*. ■

Corollary 2.5 *The group* $\mathrm{PGL}(2,F)$ *is* 3*-transitive on the points of the projective line* $\mathrm{PG}(1,F)$*; the stabiliser of three points is isomorphic to the group of inner* *automorphisms of the multiplicative group of F. In particular, if F is commutative, then* $\mathrm{PGL}(2,F)$ *is sharply 3-transitive on the points of the projective line.*

For $n > 2$, the group $\mathrm{PGL}(n,F)$ *is 2-transitive on the points of the projective space* $\mathrm{PG}(n-1,F)$.

This follows from the theorem because, in the projective plane, the hyperplanes are the points, and so no two distinct points lie in a hyperplane; while, in general, any two points are independent and can be extended to an $(n+1)$-tuple as in the theorem.

We can represent the set of points of the projective line as $\{\infty\} \cup F$, where $\infty = \langle(1,0)\rangle$ and $a = \langle(a,1)\rangle$ for $a \in F$. Then the stabiliser of the three points $\infty, 0, 1$ acts in the natural way on $F \setminus \{0,1\}$ by conjugation.

For consider the effect of the diagonal matrix aI on the point $\langle(x,1)\rangle$. This is mapped to $\langle(xa,a)\rangle$, which is the same rank 1 subspace as $\langle(a^{-1}xa,1)\rangle$; so in the new representation, aI induces the map $x \mapsto a^{-1}xa$.

In this convenient representation, the action of $\mathrm{PGL}(2,F)$ can be represented by linear fractional transformations. The matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ maps $(x,1)$ to $(xa+c, xb+d)$, which spans the same point as $((xb+d)^{-1}(xa+c),1)$ if $xb+d \neq 0$, or $(1,0)$ otherwise. Thus the transformation induced by this matrix can be written as

$$x \mapsto (xb+d)^{-1}(xa+c),$$

provided we make standard conventions about ∞ (for example, $0^{-1}a = \infty$ for $a \neq 0$ and $(\infty b+d)^{-1}(\infty a+c) = b^{-1}a$. If F is commutative, this transformation is conveniently written as a fraction:

$$x \mapsto \frac{ax+c}{bx+d}.$$

Exercise 2.1 Work out carefully all the conventions required to use the linear fractional representation of $\mathrm{PGL}(2,F)$.

Exercise 2.2 By Theorem 2.4, the order of $\mathrm{PGL}(n,q)$ is equal to the number of $(n+1)$-tuples of points of $\mathrm{PG}(n-1,q)$ for which no n lie in a hyperplane. Use this to give an alternative proof of Theorem 2.2.

Paul Cohn constructed an example of a division ring F such that all elements of $F \setminus \{0,1\}$ are conjugate in the multiplicative group of F. For a division ring F with this property, we see that $\mathrm{PGL}(2,F)$ is 4-transitive on the projective line. This is the highest degree of transitivity that can be realised in this way.

Exercise 2.3 Show that, if F is a division ring with the above property, then F has characteristic 2, and the multiplicative group of F is torsion-free and simple.

Exercise 2.4 Let F be a commutative field. Show that, for all $n \geq 2$, the group $\mathrm{PSL}(n,F)$ is 2-transitive on the points of the projective space $\mathrm{PG}(n-1,F)$; it is 3-transitive if and only if $n = 2$ and every element of F is a square.

2.2 Generation

For the rest of this section, we assume that F is a commutative field. A *transvection* of the F-vector space V is a linear map : $V \to V$ which satisfies $\mathrm{rk}(T-I) = 1$ and $(T-I)^2 = 0$. Thus, if we choose a basis such that e_1 spans the image of $T-I$ and $e_1,\ldots.e_{n-1}$ span the kernel, then T is represented by the matrix $I+U$, where U has entry 1 in the top right position and 0 elsewhere. Note that a transvection has determinant 1. The *axis* of the transvection is the hyperplane $\ker(T-I)$; this subspace is fixed elementwise by T. Dually, the *centre* of T is the image of $T-I$; every subspace containing this point is fixed by T (so that T acts trivially on the quotient space).

Thus, a transvection is a map of the form

$$x \mapsto x + (xf)a,$$

where $a \in V$ and $f \in V^*$ satisfy $af = 0$ (that is, $f \in a^\dagger$). Its centre and axis are $\langle a \rangle$ and $\ker(f)$ respectively.

The transformation of projective space induced by a transvection is called an *elation*. The matrix form given earlier shows that all elations lie in $\mathrm{PSL}(n,F)$.

Theorem 2.6 *For any $n \geq 2$ and commutative field F, the group* $\mathrm{PSL}(n,F)$ *is generated by the elations.*

Proof We use induction on n.

Consider the case $n = 2$. The elations fixing a specified point, together with the identity, form a group which acts regularly on the remaining points. (In the linear fractional representation, this elation group is

$$\{x \mapsto x + a : a \in F\},$$

fixing ∞.) Hence the group G generated by the elations is 2-transitive. So it is enough to show that the stabiliser of the two points ∞ and 0 in G is the same as in $\mathrm{PSL}(2,F)$, namely

$$\{x \mapsto a^2 x : a \in F, a \neq 0\}.$$

Given $a \in F$, $a \neq 0$, we have

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ a-1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -a^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ a-a^2 & 1 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix},$$

and the last matrix induces the linear fractional map $x \mapsto ax/a^{-1} = a^2x$, as required.

(The proof shows that two elation groups, with centres ∞ and 0, suffice to generate $\mathrm{PSL}(2,F)$.)

Now for the general case, we assume that $\mathrm{PSL}(n-1,F)$ is generated by elations. Let G be the subgroup of $\mathrm{PSL}(n,F)$ generated by elations. First, we observe that G is transitive; for, given any two points p_1 and p_2, there is an elation on the line $\langle p_1, p_2 \rangle$ carrying p_1 to p_2, which is induced by an elation on the whole space (acting trivially on a complement to the line). So it is enough to show that the stabiliser of a point p is generated by elations. Take an element $g \in \mathrm{PSL}(n,F)$ fixing p.

By induction, G_p induces at least the group $\mathrm{PSL}(n-1,F)$ on the quotient space V/p. So, multiplying g by a suitable product of elations, we may assume that g induces an element on V/p which is diagonal, with all but one of its diagonal elements equal to 1. In other words, we can assume that g has the form

$$\begin{pmatrix} \lambda & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ x_1 & x_2 & \dots & x_{n-1} & \lambda^{-1} \end{pmatrix}.$$

By further multiplication by elations, we may assume that $x_1 = \ldots = x_{n-1} = 0$. Now the result follows from the matrix calculation given in the case $n = 2$.

Exercise 2.5 A *homology* is an element of $\mathrm{PGL}(n,F)$ which fixes a hyperplane pointwise and also fixes a point not in this hyperplane. Thus, a homology is represented in a suitable basis by a diagonal matrix with all its diagonal entries except one equal to 1.

(a) Find two homologies whose product is an elation.

(b) Prove that $\mathrm{PGL}(n,F)$ is generated by homologies.

2.3 Iwasawa's Lemma

Let G be a permutation group on a set Ω: this means that G is a subgroup of the symmetric group on Ω. Iwasawa's Lemma gives a criterion for G to be simple. We will use this to prove the simplicity of $\mathrm{PSL}(n,F)$ and various other classical groups.

Recall that G is *primitive* on Ω if it is transitive and there is no non-trivial equivalence relation on Ω which is G-invariant: equivalently, if the stabiliser G_α of a point $\alpha \in \Omega$ is a maximal subgroup of G. Any 2-transitive group is primitive.

Iwasawa's Lemma is the following.

Theorem 2.7 *Let G be primitive on Ω. Suppose that there is an abelian normal subgroup A of G_α with the property that the conjugates of A generate G. Then any non-trivial normal subgroup of G contains G'. In particular, if $G = G'$, then G is simple.*

Proof Suppose that N is a non-trivial normal subgroup of G. Then $N \not\leq G_\alpha$ for some α. Since G_α is a maximal subgroup of G, we have $NG_\alpha = G$.

Let g be any element of G. Write $g = nh$, where $n \in N$ and $h \in G_\alpha$. Then

$$gAg^{-1} = nhAh^{-1}n^{-1} = nAn^{-1},$$

since A is normal in G_α. Since N is normal in G we have $gAg^{-1} \leq NA$. Since the conjugates of A generate G we see that $G = NA$.

Hence

$$G/N = NA/N \cong A/(A \cap N)$$

is abelian, whence $N \geq G'$, and we are done. ■

2.4 Simplicity

We now apply Iwasawa's Lemma to prove the simplicity of $\mathrm{PSL}(n,F)$. First, we consider the two exceptional cases where the group is not simple.

Recall that $\mathrm{PSL}(2,q)$ is a subgroup of the symmetric group S_{q+1}, having order $(q+1)q(q-1)/(q-1,2)$.

(a) If $q = 2$, then $\mathrm{PSL}(2,q)$ is a subgroup of S_3 of order 6, so $\mathrm{PSL}(2,2) \cong S_3$. It is not simple, having a normal subgroup of order 3.

(b) If $q = 3$, then $\mathrm{PSL}(2,q)$ is a subgroup of S_4 of order 12, so $\mathrm{PSL}(2,3) \cong A_4$. It is not simple, having a normal subgroup of order 4.

(c) For comparison, we note that, if $q = 4$, then $\mathrm{PSL}(2,q)$ is a subgroup of S_5 of order 60, so $\mathrm{PSL}(2,4) \cong A_5$. This group is simple.

Lemma 2.8 *The group* $\mathrm{PSL}(n,F)$ *is equal to its derived group if* $n > 2$ *or if* $|F| > 3$.

Proof The group $G = \mathrm{PSL}(n,F)$ acts transitively on incident point-hyperplane pairs. Each such pair defines a unique elation group. So all the elation groups are conjugate. These groups generate G. So the proof will be concluded if we can show that some elation group is contained in G'.

Suppose that $|F| > 3$. It is enough to consider $n = 2$, since we can extend all matrices in the argument below to rank n by appending a block consisting of the identity of rank $n-2$. There is an element $a \in F$ with $a^2 \neq 0,1$. We saw in the proof of Theorem 2.6 that $\mathrm{SL}(2,F)$ contains the matrix $\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$. Now

$$\begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a^{-1} & 0 \\ 0 & a \end{pmatrix} = \begin{pmatrix} 1 & (a^2-1)x \\ 0 & 1 \end{pmatrix};$$

this equation expresses any element of the corresponding transvection group as a commutator.

Finally suppose that $|F| = 2$ or 3. As above, it is enough to consider the case $n = 3$. This is easier, since we have more room to manoeuvre in three dimensions: we have

$$\begin{pmatrix} 1 & -x & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & x & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad \blacksquare$$

Lemma 2.9 *Let* Ω *be the set of points of the projective space* $\mathrm{PG}(n-1,F)$. *Then, for* $\alpha \in \Omega$, *the set of elations with centre* α, *together with the identity, forms an abelian normal subgroup of* G_α.

Proof This is more conveniently shown for the corresponding transvections in $\mathrm{SL}(n,F)$. But the transvections with centre spanned by the vector a consist of all maps $x \mapsto x + (xf)a$,, for $f \in A^\dagger$; these clearly form an abelian group isomorphic to the additive group of $a^\dagger$. $\blacksquare$

Theorem 2.10 *The group* $\mathrm{PSL}(n,F)$ *is simple if* $n > 2$ *or if* $|F| > 3$.

Proof let $G = \mathrm{PSL}(n,F)$. Then G is 2-transitive, and hence primitive, on the set Ω of points of the projective space. The group A of elations with centre α is an abelian normal subgroup of G_α, and the conjugates of A generate G (by Theorem 2.6, since every elation has a centre). Apart from the two excluded cases, $G = G'$. So G is simple, by Iwasawa's Lemma. ■

2.5 Small fields

We now have the family $\mathrm{PSL}(n,q)$, for $(n,q) \neq (2,2),(2,3)$ of finite simple groups. (The first two members are not simple: we observed that $\mathrm{PSL}(2,2) \cong S_3$ and $\mathrm{PSL}(2,3) \cong A_4$, neither of which is simple.) As is well-known, Galois showed that the alternating group A_n of degree $n \geq 5$ is simple.

Exercise 2.6 Prove that the alternating group A_n is simple for $n \geq 5$.

Some of these groups coincide:

Theorem 2.11 *(a)* $\mathrm{PSL}(2,4) \cong \mathrm{PSL}(2,5) \cong A_5$.

(b) $\mathrm{PSL}(2,7) \cong \mathrm{PSL}(3,2)$.

(c) $\mathrm{PSL}(2,9) \cong A_6$.

(d) $\mathrm{PSL}(4,2) \cong A_8$.

Proofs of these isomorphisms are outlined below. Many of the details are left as exercises. There are many other ways to proceed!

Theorem 2.12 *Let G be a simple group of order $(p+1)p(p-1)/2$, where p is a prime number greater than* 3. *Then $G \cong \mathrm{PSL}(2,p)$.*

Proof By Sylow's Theorem, the number of Sylow p-subgroups is congruent to 1 mod p and divides $(p+1)(p-1)/2$; also this number is greater than 1, since G is simple. So there are $p+1$ Sylow p-subgroups; and if P is a Sylow p-subgroup and $N = N_G(P)$, then $|N| = p(p-1)/2$.

Consider G acting as a permutation group on the set Ω of cosets of N. Let ∞ denote the coset N. Then P fixes ∞ and permutes the other p cosets regularly. So we can identify Ω with the set $\{\infty\} \cup \mathrm{GF}(p)$ such that a generator of P acts on Ω

as the permutation $x \mapsto x+1$ (fixing ∞). We see that N is permutation isomorphic to the group

$$\{x \mapsto a^2x+b : a,b \in \mathrm{GF}(p), a \neq 0\}.$$

More conveniently, elements of N can be represented as linear fractional transformations of Ω with determinant 1, since

$$a^2x+b = \frac{ax+a^{-1}b}{0x+a^{-1}}.$$

Since G is 2-transitive on Ω, N is a maximal subgroup of G, and G is generated by N and an element t interchanging ∞ and 0, which can be chosen to be an involution. If we can show that t is also represented by a linear fractional transformation with determinant 1, then G will be a subgroup of the group $\mathrm{PSL}(2,p)$ of all such transformations, and comparing orders will show that $G = \mathrm{PSL}(2,p)$.

We treat the case $p \equiv -1 \pmod 4$; the other case is a little bit trickier.

The element t must normalise the stabiliser of ∞ and 0, which is the cyclic group $C = \{x \mapsto a^2x\}$ of order $(p-1)/2$ (having two orbits of size $(p-1)/2$, consisting of the non-zero squares and the non-squares in $\mathrm{GF}(p)$). Also, t has no fixed points. For the stabiliser of three points in G is trivial, so t cannot fix more than 2 points; but the two-point stabiliser has odd order $(p-1)/2$. Thus t interchanges the two orbits of C.

There are various ways to show that t inverts C. One of them uses Burnside's Transfer Theorem. Let q be any prime divisor of $(p-1)/2$, and let Q be a Sylow q-subgroup of C (and hence of G). Clearly $N_G(Q) = C\langle t\rangle$, so t must centralise or invert Q. If t centralises Q, then $Q \leq Z(N_G(Q)$, and Burnside's Transfer Theorem implies that G has a normal q-complement, contradicting simplicity. So t inverts every Sylow subgroup of C, and thus inverts C.

Now $C\langle t\rangle$ is a dihedral group, containing $(p-1)/2$ involutions, one interchanging the point 1 with each point in the other C-orbit. We may choose t so that it interchanges 1 with -1. Then the fact that t inverts C shows that it interchanges a^2 with $-a^{-2}$ for each non-zero $a \in \mathrm{GF}(p)$. So t is the linear fractional map $x \mapsto -1/x$, and we are done. ■

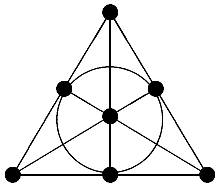
Theorem 2.11(b) follows, since $\mathrm{PSL}(3,2)$ is a simple group of order

$$(2^3-1)(2^3-2)(2^3-2^2) = 168 = (7+1)7(7-1)/2.$$

Exercise 2.7 (a) Complete the proof of the above theorem in the case $p=5$. Hence prove Theorem 2.11(a).

(b) Show that a simple group of order 60 has five Sylow 2-subgroups, and hence show that any such group is isomorphic to A_5. Give an alternative proof of Theorem 2.11(a).

Proof of Theorem 2.11(d) The simple group $\mathrm{PSL}(3,2)$ of order 168 is the group of collineations of the projective plane over $\mathrm{GF}(2)$, shown below.



Since its index in S_7 is 30, there are 30 different ways of assigning the structure of a projective plane to a given set $N = \{1,2,3,4,5,6,7\}$ of seven points; and since $\mathrm{PSL}(3,2)$, being simple, contains no odd permutations, it is contained in A_7, so these 30 planes fall into two orbits of 15 under the action of A_7.

Let Ω be one of the A_7-orbits. Each plane contains seven lines, so there $15 \times 7 = 105$ pairs (L, Π), where L is a 3-subset of N, $\Pi \in \Omega$, and L is a line of Π. Thus, each of the $\binom{7}{3} = 35$ triples is a line in exactly three of the planes in Ω.

We now define a new geometry $\mathcal{G}$ whose 'points' are the elements of Ω, and whose 'lines' are the triples of elements containing a fixed line L. Clearly, any two 'points' lie in at most one 'line', and a simple counting argument shows that in fact two 'points' lie in a unique line.

Let Π' be a plane from the other A_7-orbit. For each point $n \in N$, the three lines of Π' containing n belong to a unique plane of the set Ω. (Having chosen three lines through a point, there are just two ways to complete the projective plane, differing by an odd permutation.) In this way, each of the seven points of N gives rise to a 'point' of Ω. Moreover, the three points of a line of Π' correspond to three 'points' of a 'line' in our new geometry $\mathcal{G}$. Thus, $\mathcal{G}$ contains 'planes', each isomorphic to the projective plane $\mathrm{PG}(2,2)$.

It follows that $\mathcal{G}$ is isomorphic to $\mathrm{PG}(3,2)$. The most direct way to see this is to consider the set $A = \{0\} \cup \Omega$, and define a binary operation on A by the rules

$$\begin{array}{rl} 0+\Pi = \Pi+0 = \Pi & \text{for all } \Pi \in \Omega; \\ \Pi+\Pi = 0 & \text{for all } \Pi \in \Omega; \\ \Pi+\Pi' = \Pi'' & \text{if } \{\Pi,\Pi',\Pi''\} \text{ is a 'line'}. \end{array}$$

Then A is an elementary abelian 2-group. (The associative law follows from the fact that any three non-collinear 'points' lie in a 'plane'.) In other words, A is the additive group of a rank 4 vector space over $\mathrm{GF}(2)$, and clearly $\mathcal{G}$ is the projective geometry based on this vector space.

Now $A_7 \leq \mathrm{Aut}(\mathcal{G}) = \mathrm{PSL}(4,2)$. (The last inequality comes from the Fundamental Theorem of Projective Geometry and the fact that $\mathrm{PSL}(4,2) = \mathrm{P\Gamma L}(4,2)$ since $\mathrm{GF}(2)$ has no non-trivial scalars or automorphisms.) By calculating orders, we see that A_7 has index 8 in $\mathrm{PSL}(4,2)$. Thus, $\mathrm{PSL}(4,2)$ is a permutation group on the cosets of A_7, that is, a subgroup of S_8, and a similar calculation shows that it has index 2 in S_8. We conclude that $\mathrm{PSL}(4,2) \cong A_8$. ∎

The proof of Theorem 2.11(c) is an exercise. Two approaches are outlined below. Fill in the details.

Exercise 2.8 The field GF(9) can be represented as $\{a+b\mathrm{i} : a,b \in \mathrm{GF}(3)\}$, where $\mathrm{i}^2 = -1$. Let

$$A = \begin{pmatrix} 1 & 1+\mathrm{i} \\ 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then

$$A^3 = I, \quad B^2 = -I, \quad (AB)^5 = -I.$$

So the corresponding elements $a, b \in G = \mathrm{PSL}(2,9)$ satisfy

$$a^3 = b^2 = (ab)^5 = 1,$$

and so generate a subgroup H isomorphic to A_5. Then H has index 6 in G, and the action of G on the cosets of H shows that $G \leq S_6$. Then consideration of order shows that $G \cong A_6$.

Exercise 2.9 Let $G = A_6$, and let H be the normaliser of a Sylow 3-subgroup of G. Let G act on the 10 cosets of H. Show that H fixes one point and acts is isomorphic to the group

$$\{x \mapsto a^2x + b : a,b \in \mathrm{GF}(9), a \neq 0\}$$

on the remaining points. Choose an element outside H and, following the proof of Theorem 2.12, show that its action is linear fractional (if the fixed point is labelled as ∞). Deduce that $A_6 \leq \mathrm{PSL}(2,9)$, and by considering orders, show that equality holds.

Exercise 2.10 A *Hall subgroup* of a finite group G is a subgroup whose order and index are coprime. Philip Hall proved that a finite soluble group G has Hall subgroups of all *admissible* orders m dividing $|G|$ for which $(m, |G|/m) = 1$, and that any two Hall subgroups of the same order in a finite soluble group are conjugate.

(a) Show that $\mathrm{PSL}(2,5)$ fails to have a Hall subgroup of some admissible order.

(b) Show that $\mathrm{PSL}(2,7)$ has non-conjugate Hall subgroups of the same order.

(c) Show that $\mathrm{PSL}(2,11)$ has non-isomorphic Hall subgroups of the same order.

(d) Show that each of these groups is the smallest with the stated property.

Exercise 2.11 Show that $\mathrm{PSL}(4,2)$ and $\mathrm{PSL}(3,4)$ are non-isomorphic simple groups of the same order.

3 Polarities and forms

3.1 Sesquilinear forms

We saw in Chapter 1 that the projective space $\mathrm{PG}(n-1,F)$ is isomorphic to its dual if and only if the field F is isomorphic to its opposite. More precisely, we have the following. Let σ be an anti-automorphism of F, and V an F-vector space of rank n. A *sesquilinear form* B on V is a function $B : V \times V \to F$ which satisfies the following conditions:

(a) $B(c_1x_1 + c_2x_2, y) = c_1B(x_1, y) + c_2B(x_2, y)$, that is, B is a linear function of its first argument;

(b) $B(x, c_1y_1 + c_2y_2) = B(x, y_1)c_1^\sigma + B(x, y_2)c_2^\sigma$, that is, B is a semilinear function of its second argument, with field anti-automorphism σ.

(The word 'sesquilinear' means 'one-and-a-half'.) If σ is the identity (so that F is commutative), we say that B is a *bilinear form*.

The *left radical* of B is the subspace $\{x \in V : (\forall y \in V)B(x,) = 0\}$, and the *right radical* is the subspace $\{y \in V : (\forall x \in V)B(x, y) = 0\}$.

Exercise 3.1 (a) Prove that the left and right radicals are subspaces.

(b) Show that the left and right radicals have the same rank (if V has finite rank).

(c) Construct a bilinear form on a vector space of infinite rank such that the left radical is zero and the right radical is no-zero.

The sesquilinear form B is called *non-degenerate* if its left and right radicals are zero. (By the preceding exercise, it suffices to assume that one of the radicals is zero.)

A non-degenerate sesquilinear form induces a duality of $\mathrm{PG}(n-1,F)$ (an isomorphism from $\mathrm{PG}(n-1,F)$ to $\mathrm{PG}(n-1,F^\circ)$) as follows: for any $y \in V$, the map $x \mapsto B(x, y)$ is a linear map from V to F, that is, an element of the dual space V^* (which is a left vector space of rank n over F°); if we call this element β_y, then the map $y \mapsto \beta_y$ is a σ-semilinear bijection from V to V^*, and so induces the required duality.

Theorem 3.1 *For $n \geq 3$, any duality of* $\mathrm{PG}(n-1,F)$ *is induced in this way by a non-degenerate sesquilinear form on $V = F^n$.*

Proof By the Fundamental Theorem of Projective Geometry, a duality is induced by a σ-semilinear bijection ϕ from V to V^*, for some anti-automorphism σ. Set

$$B(x,y) = x(y\phi). \quad \blacksquare$$

We can short-circuit the passage to the dual space, and write the duality as

$$U \mapsto U^\perp = \{x \in V : B(x,y) = 0 \text{ for all } y \in U\}.$$

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 will turn out that polarities give rise to a class of geometries (the polar spaces) with properties similar to those of projective spaces, and define groups analogous to the projective groups. If a duality is not a polarity, then any collineation which respects it must commute with its square, which is a collineation; so the group we obtain will lie inside the centraliser of some element of the collineation group. So the "largest" subgroups obtained will be those preserving polarities.

A sesquilinear form B is *reflexive* if $B(x,y) = 0$ implies $B(y,x) = 0$.

Proposition 3.2 *A duality is a polarity if and only if the sesquilinear form defining it is reflexive.*

Proof B is reflexive if and only if $x \in \langle y\rangle^\perp \Rightarrow y \in \langle x\rangle^\perp$. 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 $y \in \langle x\rangle^\perp$, then $x \in \langle x\rangle^{\perp\perp} \subseteq \langle y\rangle^\perp$ (since inclusions are reversed). $\blacksquare$

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 $B(y,x) = B(x,y)^\sigma$ for all $x,y \in V$. If B is a non-zero σ-Hermitian form, then

(a) for any x, $B(x,x)$ lies in the fixed field of σ;

(b) $\sigma^2 = 1$. For every scalar c is a value of B, say $B(x,y) = c$; then

$$c^{\sigma^2} = B(x,y)^{\sigma^2} = B(y,x)^\sigma = B(x,y) = c.$$

If σ is the identity, such a form (which is bilinear) is called *symmetric*.

A bilinear form b is called *alternating* if $B(x,x) = 0$ for all $x \in V$. This implies that $B(x,y) = -B(y,x)$ for all $x,y \in V$. For

$$0 = B(x+y,x+y) = B(x,x) + B(x,y) + B(y,x) + B(y,y) = B(x,y) + B(y,x).$$

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 3.3 *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.

Proof I will give the proof just for a bilinear form. Thus, it must be proved that a non-degenerate reflexive bilinear form is either symmetric or alternating.

We have

$$B(u,v)B(u,w) - B(u,w)B(u,v) = 0$$

by commutativity; that is, using bilinearity,

$$B(u, B(u,v)w - B(u,w)v) = 0.$$

By reflexivity,

$$B(B(u,v)w - B(u,w)v, u) = 0,$$

whence bilinearity again gives

$$B(u,v)B(w,u) = B(u,w)B(v,u). \tag{1}$$

Call a vector u *good* if $B(u,v) = B(v,u) \neq 0$ for some v. By Equation (1), if u is good, then $B(u,w) = B(w,u)$ for all w. Also, if u is good and $B(u,v) \neq 0$, then v is good. But, given any two non-zero vectors u_1, u_2, there exists v with $B(u_i, v) \neq 0$ for $i = 1,2$. (For there exist v_1, v_2 with $B(u_i, v_i) \neq 0$ for $i = 1,2$, by non-degeneracy; and at least one of $v_1, v_2, v_1 + v_2$ has the required property.) So, if some vector is good, then every non-zero vector is good, and B is symmetric.

But, putting $u = w$ in Equation (1) gives

$$B(u,u)\,(B(u,v) - B(v,u)) = 0$$

for all u,v. So, if u is not good, then $B(u,u) = 0$; and, if no vector is good, then B is alternating. ■

Exercise 3.2 (a) Show that the left and right radicals of a reflexive form are equal.

(b) Assuming Theorem 3.3, prove that the assumption of non-degeneracy in the theorem can be removed.

Exercise 3.3 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 $\varepsilon = (a+1) + b\omega$.]

Exercise 3.4 Use the result of the preceding exercise to complete the proof of Theorem 3.3 in general.

[If $B(u,u) = 0$ for all u, the form B is alternating and bilinear. If not, suppose that $B(u,u) \neq 0$ and let $B(u,u)^\sigma = \lambda B(u,u)$. Choosing ε as in Exercise 3.3 and re-normalising B, show that we may assume that $\lambda = 1$, and (with this choice) that B is Hermitian.]

3.2 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 $B(x,x) \in \mathrm{Fix}(\sigma)$ for all $x \in V$. We call the form B *trace-valued* if $B(x,x) \in \mathrm{Tr}(\sigma)$ for all $x \in V$.

Exercise 3.5 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 $\mathrm{Fix}(\sigma)$.

Proposition 3.4 $\mathrm{Tr}(\sigma) = \mathrm{Fix}(\sigma)$ *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 3.5). 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 Theorem 3.3 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*. We will not pursue this here: see Tits's classification of spherical buildings.

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

(a) $q(\lambda x) = \lambda^2 f(x)$ for all $\lambda \in F$, $x \in V$;

(b) $q(x+y) = q(x) + q(y) + B(x,y)$, where B is bilinear.

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

$$\begin{aligned} B(x,y) &= q(x+y) - q(x) - q(y), \\ q(x) &= \tfrac{1}{2}B(x,x), \end{aligned}$$

the latter equation coming from the substitution $x = y$ in (b). So nothing new is obtained.

On the other hand, if the characteristic of F is 2, then B is an alternating bilinear form, and q 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 B is obtained by *polarisation* of q.

Now let B be a symmetric bilinear form over a field of characteristic 2, which is not alternating. Set $f(x) = B(x,x)$. Then we have

$$\begin{aligned} f(\lambda x) &= \lambda^2 f(x), \\ f(x+y) &= f(x) + f(y), \end{aligned}$$

since $B(x,y) + B(y,x) = 0$. 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 q is *non-degenerate* if

$$q(x) = 0 \ \& \ (\forall y \in V)B(x,y) = 0 \quad \Rightarrow \quad x = 0,$$

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

If the characteristic is not 2, then non-degeneracy of the quadratic form and of the bilinear form are equivalent conditions.

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

$$\begin{aligned} q(x+y) &= q(x)+q(y), \\ q(\lambda x) &= \lambda^2 q(x). \end{aligned}$$

As above, f is very nearly semilinear.

The field F is called *perfect* if every element is a square. If F is perfect, then the map $x \mapsto x^2$ is onto, and hence an automorphism of F; so q is indeed semilinear, and its kernel is a hyperplane of W. We conclude:

Theorem 3.5 *Let q 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.

Exercise 3.6 Let B be an alternating bilinear form on a vector space V over a field F of characteristic 2. Let $(v_i : i \in I)$ be a basis for V, and $(c_i : i \in I)$ any function from I to F. Show that there is a unique quadratic form q with the properties that $q(v_i) = c_i$ for every $i \in I$, and q polarises to B.

Exercise 3.7 (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.

Exercise 3.8 Show that finite fields of characteristic 2 are perfect.

Exercise 3.9 Let B be a σ-Hermitian form on a vector space V over F, where σ is not the identity. Set $f(x) = B(x,x)$. 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 $\mathrm{Tr}(B)(x,y) = B(x,y) + B(x,y)^\sigma$. Show further that $\mathrm{Tr}(B)$ is non-degenerate if and only if B is.

3.3 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 trace-valued σ-Hermitian form B, where σ is not the identity; or a non-singular quadratic form q. In the third case, we let B be the bilinear form obtained by polarising q; then B is alternating or symmetric according as the characteristic is or is not 2, but B may be degenerate. We also let f denote the function q. In the other two cases, we define a function $f : V \to F$ by $f(x) = B(x,x)$ — this is identically zero if b is alternating. See Exercise 3.10 for the Hermitian case.

Such a pair (V,B) or (V,q) will be called a *formed space*.

Exercise 3.10 Let B be a σ-Hermitian form on a vector space V over F, where σ is not the identity. Set $f(x) = B(x,x)$. 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 $\mathrm{Tr}(B)(x,y) = B(x,y) + B(x,y)^{\sigma}$. Show further that $\mathrm{Tr}(b)$ is non-degenerate if and only if B is.

We say that V is *anisotropic* if $f(x) \neq 0$ for all $x \neq 0$. Also, V is a *hyperbolic plane* if it is spanned by vectors v and w with $f(v) = f(w) = 0$ and $B(v,w) = 1$. (The vectors v and w are linearly independent, so V has rank 2.)

Theorem 3.6 *A non-degenerate formed space 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, since V cannot contain a hyperbolic plane. So suppose that V contains a vector $v \neq 0$ with $f(v) = 0$.

We claim that there is a vector w with $B(v,w) \neq 0$. 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 v is in the radical of B; but v is a singular vector, contradicting the non-degeneracy of f.

Multiplying w by a non-zero constant, we may assume that $B(v,w) = 1$.

Now, for any value of λ, we have $B(v, w - \lambda v) = 1$. We wish to choose λ so that $f(w - \lambda v) = 0$; then v and w will span a hyperbolic line. Now we distinguish cases.

(a) If B is alternating, then any value of λ works.

(b) If B is Hermitian, we have

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

and, since B is trace-valued, there exists λ with $\mathrm{Tr}(\lambda)=f(w)$.

(c) Finally, if $f=q$ is quadratic, we have

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

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

Now let W_1 be the hyperbolic line $\langle v, w-\lambda v\rangle$, 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. Now the existence of the decomposition follows by induction.

The uniqueness of the decomposition will be proved later, as a consequence of Witt's Lemma (Theorem 3.15). ■

The number r of hyperbolic lines is called the *polar rank* of V, and (the isomorphism type of) U is called the *germ* of V.

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 3.7 *The only anisotropic space carrying an alternating bilinear form is the zero space.*

In combination with Theorem 3.6, this shows that a space carrying a non-degenerate alternating bilinear form is a direct sum of hyperbolic planes.

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 3.8 *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 3.9 *(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 below).

(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 3.3).

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. ■

Exercise 3.11 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 planes).

Exercise 3.12 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.

3.4 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 $f(x) = B(x,x)$, 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 *B-flat subspace*, and one on which f vanishes identically is called a *f-flat subspace*. (Note: these terms are not standard; in the literature, such spaces are called *totally isotropic* (t.i.) and *totally singular* (t.s.) respectively.) The unqualified term *flat subspace* will mean a B-flat subspace in the symplectic or unitary case, and a q-flat subspace in the orthogonal case.

The *polar space* associated with a vector space carrying a form is the geometry whose flats are the flat subspaces (in the above sense). 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 flat 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 rank as the number of hyperbolic planes in the decomposition of the space.) We use the terms *point, line, plane*, etc., just as for projective spaces.

Polar spaces bear the same relation to formed spaces as projective spaces do to vector spaces.

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

(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 planes 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 $r-1$.

(P1) is clear since a subspace of a flat subspace is itself flat. (P2) is also clear. To prove (P3), let $p = \langle y \rangle$. The function $x \mapsto B(x,y)$ 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 flat if and only if $q \in K$; and the union of all such flat lines is a flat space $W = \langle K, y \rangle$, such that $W \cap U = K$, as required.

Finally, to prove (P4), we use the hyperbolic-anisotropic decomposition again. If $L_1, \ldots, L_r$ are the hyperbolic planes, and x_i, y_i are the distinguished spanning vectors in L_i, then the required flats are $\langle x_1, \ldots, x_r \rangle$ and $\langle y_1, \ldots, y_r \rangle$.

The significance of the geometric properties (P1)–(P4) lies in the major result of Veldkamp and Tits which determines all the geometries of rank at least 3 which satisfy them. All these geometries are polar spaces (as we have defined them) or slight generalisations, together with a couple of exceptions of rank 3. In particular, the following theorem holds:

Theorem 3.10 *A finite geometry satisfying (P1)–(P4) with $r \geq 3$ is a polar space.*

Exercise 3.13 Let $P = \mathrm{PG}(3,F)$ for some (not necessarily commutative) division ring F. Construct a new geometry Γ as follows:

(a) the 'points' of Γ are the lines of P;

(b) the 'lines' of Γ are the plane pencils in P (consisting of all lines lying in a plane Π and containing a point p of Π);

(c) the 'planes' of Γ are of two types: the pencils (consisting of all the lines through a point) and the dual planes (consisting of all the lines in a plane).

Prove that Γ satisfies (P1)–(P4) with $r = 3$.

Prove that, if F is not isomorphic to its opposite, then Γ contains non-isomorphic planes.

(We will see later that, if F is commutative, then Γ is an orthogonal polar space.)

Exercise 3.14 Prove the *Buekenhout–Shult property* of the geometry of points and lines in a polar space: if p is a point not lying on a line L, then p is collinear with one or all points of L.

You should prove this both from the analytic description of polar spaces, and using (P1)–(P4).

In a polar space Γ, given 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. 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 plane.)

Exercise 3.15 Prove the above assertions.

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 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*, the orthogonal space which is a direct sum of hyperbolic planes (that is, having germ 0). The quadratic form defining this space can be taken to be $x_1x_2+x_3x_4+\ldots+x_{2r-1}x_{2r}$.

Proposition 3.11 *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 flat 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 Figure 1. This is the so-called "ruled quadric", familiar from models such as wastepaper baskets.

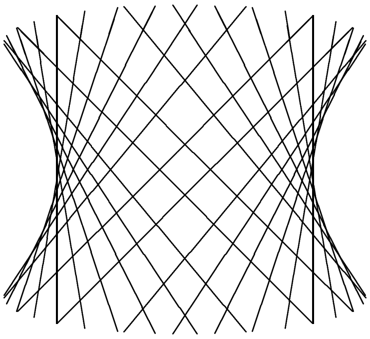


Figure 1: A ruled quadric

Exercise 3.16 Show that Proposition 3.11 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.

3.5 Finite polar spaces

The classification of finite classical polar spaces was achieved by Theorem 3.6. 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, and δ the vector space rank of the germ; the rank n of the space is given by $n = 2r + \delta$. 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	δ	ε
Symplectic	0	0
Unitary	0	$-\frac{1}{2}$
Unitary	1	$\frac{1}{2}$
Orthogonal	0	-1
Orthogonal	1	0
Orthogonal	2	1

Table 1: Finite polar spaces

Theorem 3.12 *The number of points in a finite polar space of rank 1 is* $q^{1+\varepsilon} + 1$*, where* ε *is given in Table 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 3.13 *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.

Claim 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.

Claim 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 Claim 2 and correctly reduces to $q^{1+\varepsilon}+1$ when $r=1$. ■

Theorem 3.14 *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. We will do this for the associated classical groups at the end of the next section.

3.6 Witt's Lemma

Let V be a formed space, with sesquilinear form B and (if appropriate) quadratic form q. An *isometry* of V is a linear map $g : V \to V$ which satisfies $B(xg, yg) = B(x, y)$ for all $x, y \in V$, and (if appropriate) $q(xg) = q(x)$ for all $x \in V$. (Note that, in the case of a quadratic form, the second condition implies the first.)

The set of all isometries of V forms a group, the *isometry group* of V. This group is our object of study for the next few sections.

More generally, if V and W are formed spaces of the same type, an isometry from V to W is a linear map from V to W satisfying the conditions listed above.

Exercise 3.17 Let V be a (not necessarily non-degenerate) formed space of symplectic or Hermitian type, with radical $V^\perp$. Prove that the natural map from V to $V/V^\perp$ is an isometry.

The purpose of this subsection is to prove *Witt's Lemma*, a transitivity assertion about the isometry group of a formed space.

Theorem 3.15 *Suppose that U_1 and U_2 are subspaces of the formed space V, and $h : U_1 \to U_2$ is an isometry. Then there is an isometry g of V which extends h if and only if $(U_1 \cap V^\perp)h = U_2 \cap V^\perp$.*

In particular, if $V^\perp = 0$, then any isometry between subspaces of V extends to an isometry of V.

Proof Assume that $h : U_1 \to U_2$ is an isometry. Clearly, if h is the restriction of an isometry g of V, then $V^\perp g = V^\perp$, and so

$$(U_1 \cap V^\perp)h = (U_1 \cap V^\perp)g = U_1 g \cap V^\perp g = U_2 \cap V^\perp.$$

We have to prove the converse.

First we show that we may assume that U_1 and U_2 contain $V^\perp$. Suppose not. Choose a subspace W of $V^\perp$ which is a complement to both $U_1 \cap V^\perp$ and $U_2 \cap V^\perp$ (see Exercise 3.18), and extend h to $U_1 \oplus W$ by the identity map on W. This is easily checked to be an isometry to $U_2 \oplus W$.

The proof is by induction on $\mathrm{rk}(U_1/V^\perp)$. If $U_1 = V_\perp = U_2$, then choose any complement W for $V^\perp$ in V and extend h by the identity on W. So the base step of the induction is proved. Assume that the conclusion of Witt's Lemma holds for V', U_1', U_2', h' whenever $\mathrm{rk}(U_1'/(V')^\perp) < \mathrm{rk}(U_1/V^\perp)$.

Let H be a hyperplane of U_1 containing $V^\perp$. Then the restriction f' of f to H has an extension to an isometry g' of V. Now it is enough to show that $h(g')^{-1}$ extends to an isometry; in other words, we may assume that h is the identity on H. Moreover, the conclusion is clear if h is the identity on U_1; so suppose not. Then $\ker(h-1) = H$, and so the image of $h-1$ is a rank 1 subspace P of U_1.

Since h is an isometry, for all $x, y \in U_1$ we have

$$\begin{aligned} B(xh, yh-y) &= B(xh, yh) - B(xh, y) \\ &= B(x, y) - B(xh, y) \\ &= B(x - xh, y). \end{aligned}$$

So, if $y \in H$, then any vector $xh - x$ of P is orthogonal to y; that is, $H \le P^\perp$.

Now suppose that $P \not\le U_1^\perp$. Then $U_1 \cap P^\perp = U_2 \cap P^\perp = H$. If W is a complement to H in $P^\perp$, then we can extend h by the identity on W to obtain the required isometry. So we may assume further that $U_1, U_2 \le P^\perp$. In particular, $P \le P^\perp$.

Next we show that we may assume that $U_1 = U_2 = P^\perp$. Suppose first that $U_1 \ne U_2$. If $U_i = \langle H, u_i \rangle$ for $i = 1, 2$, let W_0 be a complement for $U_1 + U_2$ in $P^\perp$, and $W = \langle W_0, u_1 + u_2 \rangle$; then h can be extended by the identity on W to an isometry on $P^\perp$. If $U_1 = U_2$, take any complement W to U_1 in $P^\perp$. In either case, the extension is an isometry of $P^\perp$ which acts as the identity on a hyperplane H' of $P^\perp$ containing H. So we may replace U_1, U_2, H by $P^\perp, P^\perp, H'$.

Let $P = \langle x \rangle$ and let $x = uh - u$ for some $u \in U_1$. We have $B(x, x) = 0$. In the orthogonal case, we have

$$q(x) = q(uh-u) = q(uh) + q(u) - B(uh, u) = 2q(u) - B(u, u) = 0.$$

(We have $B(uh,u) = B(u,u)$ because $B(uh-u,u) = 0$.) So P is flat, and there is a hyperbolic plane $\langle u,v\rangle$, with $v \notin P^{\perp}$. Our job is to extend h to the vector v.

To achieve this, we show first that there is a vector v' such that $\langle uh,v'\rangle^{\perp} = \langle u,v\rangle^{\perp}$. This holds because $\langle u,v\rangle^{\perp}$ is a hyperplane in $\langle uh\rangle^{\perp}$ not containing $V^{\perp}$.

Next, we observe that $\langle uh,v'\rangle$ is a hyperbolic plane, so we can choose a vector v'' such that $B(uh,v'') = 1$ and (if relevant) $Q(v'') = 0$.

Finally, we observe that by extending h to map v to v'' we obtain the required isometry of V.

Exercise 3.18 Let U_1 and U_2 be subspaces of a vector space V having the same rank. Show that there is a subspace W of V which is a complement for both U_1 and U_2.

Corollary 3.16 *(a) The ranks of maximal flat subspaces of a formed space are all equal.*

(b) The Witt rank and isometry type of the germ of a non-degenerate formed space are invariants.

Proof (a) Let U_1 and U_2 be maximal flat subspaces. Then both U_1 and U_2 contains $V^{\perp}$. If $\mathrm{rk}(U_1) < \mathrm{rk}(U_2)$, there is an isometry h from U_1 into U_2. If g is the extension of h to V, then the image of U_2 under g^{-1} is a flat subspace properly containing U_1, contradicting maximality.

(b) The result is clear if V is anisotropic. Otherwise, let U_1 and U_2 be hyperbolic planes. Then U_1 and U_2 are isometric and are disjoint from $V^{\perp}$. An isometry of V carrying U_1 to U_2 takes $U_1^{\perp}$ to $U_2^{\perp}$. Then the result follows by induction. ∎

Theorem 3.17 *Let V_r be a non-degenerate formed space with polar rank r and germ W over* GF(q). *Let G_r be the isometry group of V_r. Then*

$$\begin{aligned} |G_r| &= \left(\prod_{i=1}^{r}(q^i-1)(q^{i+\varepsilon}+1)q^{2i-1+\varepsilon}\right)|G_0| \\ &= q^{r(r+\varepsilon)}\left(\prod_{i=1}^{r}(q^i-1)(q^{i+\varepsilon}+1)\right)|G_0|, \end{aligned}$$

where $|G_0|$ is given by the following table:

Type	δ	ε	$\lvert G_0\rvert$
Symplectic	0	0	1
Unitary	0	$-\frac{1}{2}$	1
Unitary	1	$\frac{1}{2}$	$q^{1/2}+1$
Orthogonal	0	-1	1
Orthogonal	1	0	$\begin{cases} 2 & (q \text{ odd}) \\ 1 & (q \text{ even}) \end{cases}$
Orthogonal	2	1	$2(q+1)$

Proof By Theorem 3.13, the number of choices of a vector x spanning a flat subspace is $(q^r-1)(q^{r+\varepsilon}+1)$. Then the number of choices of a vector y spanning a flat subspace and having inner product 1 with x is $q^{2r-1+\varepsilon}$. Then x and y span a hyperbolic plane. Now Witt's Lemma shows that G_r acts transitively on such pairs, and the stabiliser of such a pair is G_{r-1}, by the inductive principle.

In the cases where $\delta = 0$, G_0 is the trivial group on a vector space of rank 0. In the unitary case with $\delta = 1$, G_0 preserves the Hermitian form $xx^{q^{1/2}}$, so consists of multiplication by $(q^{1/2}+1)$st roots of unity. In the orthogonal case with $\delta = 1$, G_0 preserves the quadratic form x^2, and so consists of multiplication by ± 1 only. Finally, consider the orthogonal case with $\delta = 2$. Here we can represent the quadratic form as the norm from $\mathrm{GF}(q^2)$ to $\mathrm{GF}(q)$, that is, $N(x) = x^{q+1}$. The $\mathrm{GF}(q)$-linear maps which preserve this form a dihedral group of order $2(q+1)$: the cyclic group is generated by the $(q+1)$st roots of unity in $\mathrm{GF}(q^2)$, which is inverted by the non-trivial field automorphism over $\mathrm{GF}(q)$ (since, if $x^{q+1} = 1$, then $x^q = x^{-1}$).

4 Symplectic groups

In this and the next two sections, we begin the study of the groups preserving reflexive sesquilinear forms or quadratic forms. We begin with the symplectic groups, associated with non-degenerate alternating bilinear forms.

4.1 The Pfaffian

The determinant of a skew-symmetric matrix is a square. This can be seen in small cases by direct calculation:

$$\begin{aligned}
\det\begin{pmatrix} 0 & a_{12} \\ -a_{12} & 0 \end{pmatrix} &= a_{12}^2, \\
\det\begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{pmatrix} &= (a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23})^2.
\end{aligned}$$

Theorem 4.1 *(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

$$PAP^\top = \operatorname{diag}\left(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right),$$

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

$$(1\,2)(3\,4)\cdots(2n-1\,2n),$$

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}). ∎

Theorem 4.2 *If A is a skew-symmetric matrix and P any invertible matrix, then*

$$\mathrm{Pf}(PAP^\top) = \det(P) \cdot \mathrm{Pf}(A).$$

Proof 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. For this, it suffices to consider the 'standard' skew-symmetric matrix

$$A = \operatorname{diag}\left(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right),$$

since all non-singular skew-symmetric matrices are equivalent. In this case, 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 $+1$. ∎

Exercise 4.1 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

$$\mathrm{Pf}(A)=\sum_{F\in\mathcal{F}_n}(-1)^{\chi(F)}\prod_{(i,j)\in F}a_{ij}.$$

4.2 The symplectic groups

The *symplectic group* $\mathrm{Sp}(2n,F)$ is the isometry group of a non-degenerate alternating bilinear form on a vector space of rank $2n$ over F. (We have seen that any two such forms are equivalent up to invertible linear transformation of the variables; so we have defined the symplectic group uniquely up to conjugacy in $\mathrm{GL}(2n,F)$.) Alternatively, it consists of the $2n\times 2n$ matrices P satisfying $P^\top AP=A$, where A is a fixed invertible skew-symmetric matrix. If necessary, we can take for definiteness either

$$A=\begin{pmatrix}O_n & I_n\\ -I_n & O_n\end{pmatrix}$$

or

$$A=\mathrm{diag}\left(\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix},\ldots,\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\right).$$

The *projective symplectic group* $\mathrm{PSp}(2n,F)$ is the group induced on the set of points of $\mathrm{PG}(2n-1,F)$ by $\mathrm{Sp}(2n,F)$. It is isomorphic to the factor group $\mathrm{Sp}(2n,F)/(Sp(2n,F)\cap Z)$, where Z is the group of non-zero scalar matrices.

Proposition 4.3 *(a)* $\mathrm{Sp}(2n,F)$ *is a subgroup of* $\mathrm{SL}(2n,F)$.

(b) $\mathrm{PSp}(2n,F)\cong\mathrm{Sp}(2n,F)/\{\pm I\}$.

Proof (a) If $P\in\mathrm{Sp}(2n,F)$, then $\mathrm{Pf}(A)=\mathrm{Pf}(PAP^\top)=\det(P)\,\mathrm{Pf}(A)$, so $\det(P)=1$.

(b) If $(cI)A(cI)=A$, then $c^2=1$, so $c=\pm1$. ■

From Theorem 3.17, we have:

Proposition 4.4

$$|\,\mathrm{Sp}(2n,q)\,| = \prod_{i=1}^{n}(q^{2i}-1)q^{2i-1} = q^{n^2}\prod_{i=1}^{n}(q^{2i}-1). \quad \blacksquare$$

The next result shows that we get nothing new in the case $2n = 2$.

Proposition 4.5 $\mathrm{Sp}(2,F) \cong \mathrm{SL}(2,F)$ *and* $\mathrm{PSp}(2,F) \cong \mathrm{PSL}(2,F)$.

Proof We show that there is a non-degenerate bilinear form on F^2 preserved by $\mathrm{SL}(2,F)$. The form B is given by

$$B(x,y) = \det\begin{pmatrix} x \\ y \end{pmatrix}$$

for all $x,y \in F^2$, where $\begin{pmatrix} x \\ y \end{pmatrix}$ is the matrix with rows x and y. This is obviously a symplectic form. For any linear map $P : F^2 \to F^2$, we have

$$\begin{pmatrix} xP \\ yP \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} P,$$

whence

$$B(xP,yP) = \det\begin{pmatrix} xP \\ yP \end{pmatrix} = B(x,y)\det(P),$$

and so all elements of $\mathrm{SL}(2,F)$ preserve B, as required.

The second assertion follows on factoring out the group of non-zero scalar matrices of determinant 1, that is, $\{\pm I\}$. $\blacksquare$

In particular, $\mathrm{PSp}(2,F)$ is simple if and only if $|F| > 3$.

There is one further example of a non-simple symplectic group:

Proposition 4.6 $\mathrm{PSp}(4,2) \cong S_6$.

Proof Let $F = \mathrm{GF}(2)$ and $V = F^6$. On V define the "standard inner product"

$$x \cdot y = \sum_{i=1}^{6} x_i y_i$$

(evaluated in F). Let j denote the all-1 vector. Then

$$x \cdot x = x \cdot j$$

for all $x \in X$, so on the rank 5 subspace $j^\perp$, the inner product induces an alternating bilinear form. This form is degenerate — indeed, by definition, its radical contains j — but it induces a non-degenerate symplectic form B on the rank 4 space $j^\perp/\langle j\rangle$. Clearly any permutation of the six coordinates induces an isometry of B. So $S_6 \leq \mathrm{Sp}(4,2) = \mathrm{PSp}(4,2)$. Since

$$|S_6| = 6! = 15 \cdot 8 \cdot 3 \cdot 2 = |\mathrm{Sp}(4,2)|,$$

the result is proved. ■

4.3 Generation and simplicity

This subsection follows the pattern used for $\mathrm{PSL}(n,F)$. We show that $\mathrm{Sp}(2n,F)$ is generated by transvections, that it is equal to its derived group, and that $\mathrm{PSp}(2n,F)$ is simple, for $n \geq 2$, with the exception (noted above) of $\mathrm{PSp}(4,2)$.

Let B be a symplectic form. Which transvections preserve B? Consider the transvection $x \mapsto x + (xf)a$, where $a \in V$, $f \in V^*$, and $af = 0$. We have

$$B(x+(xf)a, y+(yf)a) = B(x,y) + (xf)B(a,y) - (yf)B(a,x).$$

So B is preserved if and only if $(xf)B(a,y) = (yf)B(a,x)$ for all $x,y \in V$. We claim that this entails $xf = \lambda B(a,x)$ for all x, for some scalar λ. For we can choose x with $B(a,x) \neq 0$, and define $\lambda = (xf)/B(a,x)$; then the above equation shows that $yf = \lambda B(a,y)$ for all y.

Thus, a *symplectic transvection* (one which preserves the symplectic form) can be written as

$$x \mapsto x + \lambda B(x,a)a$$

for a fixed vector $a \in V$. Note that its centre and axis correspond under the symplectic polarity; that is, its axis is $a^\perp = \{x : B(x,a) = 0\}$.

Lemma 4.7 *For $r \geq 1$, the group $\mathrm{PSp}(2r,F)$ acts primitively on the points of $\mathrm{PG}(2r-1,F)$.*

Proof For $r = 1$ we know that the action is 2-transitive, and so is certainly primitive. So suppose that $r \geq 2$.

Every point of $\mathrm{PG}(2r-1,F)$ is flat, so by Witt's Lemma, the symplectic group acts transitively. Moreover, any pair of distinct points spans either a flat subspace or a hyperbolic plane. Again, Witt's Lemma shows that the group is transitive on the pairs of each type. (In other words $G = \mathrm{PSp}(2r,F)$ has three orbits on ordered pairs of points, including the diagonal orbit

$$\Delta = \{(p,p) : p \in \mathrm{PG}(2r-1,F)\};$$

we say that $\mathrm{PSp}(2r,F)$ is a *rank* 3 *permutation group* on $\mathrm{PG}(2r-1,F)$.)

Now a non-trivial equivalence relation preserved by G would have to consist of the diagonal and one other orbit. So to finish the proof, we must show:

(a) if $B(x,y) = 0$, then there exists z such that $B(x,z), B(y,z) \neq 0$;

(b) if $B(x,y) \neq 0$, then there exists z such that $B(x,z) = B(y,z) \neq 0$.

This is a simple exercise. ■

Exercise 4.2 Prove (a) and (b) above.

Lemma 4.8 *For $r \geq 1$, the group $\mathrm{Sp}(2r,F)$ is generated by symplectic transvections.*

Proof The proof is by induction by r, the case $r = 1$ having been settled earlier (Theorem 2.6).

First we show that the group H generated by transvections is transitive on the non-zero vectors. Let $u, v \neq 0$. If $B(u,v) \neq 0$, then the symplectic transvection

$$x \mapsto x + \frac{B(x, v-u)}{B(u,v)}(v-u)$$

carries u to v. If $B(u,v) = 0$, choose w such that $B(u,w), B(v,w) \neq 0$ (by (a) of the preceding lemma) and map u to w to v in two steps.

Now it is enough to show that any symplectic transformation g fixing a non-zero vector u is a product of symplectic transvections. By induction, since the stabiliser of u is the symplectic group on $u^\top/\langle u \rangle$, we may assume that g acts trivially on this quotient; but then g is itself a symplectic transvection. ■

Lemma 4.9 *For $r \geq 3$, and for $r = 2$ and $F \neq \mathrm{GF}(2)$, the group $\mathrm{PSp}(2r,F)$ is equal to its derived group.*

Proof If $F \neq \mathrm{GF}(2), \mathrm{GF}(3)$, we know from Lemma 2.8 that any element inducing a transvection on a hyperbolic plane and the identity on the complement is a commutator, so the result follows. The same argument completes the proof provided that we can show that it holds for $\mathrm{PSp}(6,2)$ and $\mathrm{PSp}(4,3)$.

In order to handle these two groups, we first develop some notation which can be more generally applied. For convenience we re-order the rows and columns of the 'standard skew-symmetric matrix' so that it has the form

$$J = \begin{pmatrix} O & I \\ -I & O \end{pmatrix},$$

where O and I are the $r \times r$ zero and identity matrices. (In other words, the ith and $(i+r)$th basis vectors form a hyperbolic pair, for $i = 1, \ldots, r$.) Now a matrix C belongs to the symplectic group if and only if $C^\top JC = J$. In particular, we find that

(a) for all invertible $r \times r$ matrices A, we have

$$\begin{pmatrix} A^{-1} & O \\ O & A^\top \end{pmatrix} \in \mathrm{Sp}(2r, F);$$

(b) for all *symmetric* $r \times r$ matrices B, we have

$$\begin{pmatrix} I & B \\ O & I \end{pmatrix} \in \mathrm{Sp}(2r, F).$$

Now straightforward calculation shows that the commutator of the two matrices in (a) and (b) is equal to

$$\begin{pmatrix} I & B - ABA^\top \\ O & I \end{pmatrix},$$

and it suffices to choose A and B such that A is invertible, B is symmetric, and $B - ABA^\top$ has rank 1.

The following choices work:

(a) $r = 2, F = \mathrm{GF}(3), A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$;

(b) $r = 3, F = \mathrm{GF}(2), A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. ■

Theorem 4.10 *The group* $\mathrm{PSp}(2r,F)$ *is simple for* $r \geq 1$*, except for the cases* $(r,F) = (1,\mathrm{GF}(2))$*,* $(1,\mathrm{GF}(3))$*, and* $(2,\mathrm{GF}(2))$*.*

Proof We now have all the ingredients for Iwasawa's Lemma (Theorem 2.7), which immediately yields the conclusion. ■

As we have seen, the exceptions in the theorem are genuine.

Exercise 4.3 Show that $\mathrm{PSp}(4,3)$ is a finite simple group which has no 2-transitive action.

The only positive integers n such that $n(n-1)$ divides $|\mathrm{PSp}(4,3)|$ are $n = 2,3,4,5,6,9,10,16,81$. It suffices to show that the group has no 2-transitive action of any of these degrees. Most are straightforward but $n = 16$ and $n = 81$ require some effort.

(It is known that $\mathrm{PSp}(4,3)$ is the smallest non-abelian finite simple group with this property.)

4.4 A technical result

The result in this section will be needed at one point in our discussion of the unitary groups. It is a method of recognising the groups $\mathrm{PSp}(4,F)$ geometrically.

Consider the polar space associated with $\mathrm{PSp}(4,F)$. Its points are all the points of the projective space $\mathrm{PG}(3,F)$, and its lines are the flat lines (those on which the symplectic form vanishes). We call them F-lines for brevity. Note that the F-lines through a point p of the porojective space form the plane pencil consisting of all the lines through p in the plane $p^\perp$, while dually the F-lines in a plane Π are all those lines of Π containing the point $\Pi^\perp$. Now two points are prthogonal if and only if they line on an F-line.

The geometry of F-lines has the following property:

(a) Given an F-line L and a point p not on L, there is a unique point $q \in L$ such that pq is an F-line.

(The point q is $p^\perp \cap L$.) A geometry with this property (in which two points lie on at most one line) is called a *generalised quadrangle*.

Exercise 4.4 Show that a geometry satisfying the polar space axioms with $r = 2$ is a generalised quadrangle, and conversely.

We wish to recognise, within the geometry, the remaining lines of the projective space. These correspond to hyperbolic planes in the vector space, so we will call them H-lines. Note that the points of a H-line are pairwise non-orthogonal.

We observe that, given any two points p,q not lying on an F-line, the set

$$\{r : pr \text{ and } qr \text{ are F-lines}\}$$

is the set of points of $\{p,q\}^\perp$, and hence is the H-line containing p and q. This definition works in any generalized quadrangle, but in this case we have more:

(b) Any two points lie on either a unique F-line or a unique H-line.

(c) The F-lines and H-lines within a set $p^\perp$ form a projective plane.

(d) Any three non-collinear points lie in a unique set $p^\perp$.

Exercise 4.5 Prove conditions (b)–(d).

Conditions (a)–(d) guarantee that the geometry of F-lines and H-lines is a projective space, hence is isomorphic to $\mathrm{PG}(3,F)$ for some (possibly non-commutative) field F. Then the correspondence $p \leftrightarrow p^\perp$ is a polarity of the projective space, such that each point is incident with the corresponding plane. By the Fundamental Theorem of Projective Geometry, this polarity is induced by a symplectic form B on a vector space V of rank 4 over F (which is necessarily commutative).

Hence, again by the FTPG, the automorphism group of the geometry is induced by the group of semilinear transformations of V which preserve the set of pairs $\{(x,y) : B(x,y) = 0\}$. These transformations are composites of linear transformations preserving B up to a scalar factor, and field automorphisms. It follwos that, if $F \neq \mathrm{GF}(2)$, the automorphism group of the geometry has a unique minimal normal subgroup, which is isomorphic to $\mathrm{PSp}(4,F)$.

5 Unitary groups

In this section we analyse the unitary groups in a similar way to the treatment of the symplectic groups in the last section. Note that the treatment here applies only to the isometry groups of Hermitian forms which are not anisotropic. So in particular, the Lie groups $\mathrm{SU}(n)$ over the complex numbers are not included.

Let V be a vector space over F, σ an automorphism of F of order 2, and B a non-degenerate σ-Hermitian form on V (that is, $B(y,x) = B(x,y)^\sigma$ for all $x, y \in V$). It is often convenient to denote c^σ by $\overline{c}$, for any element $c \in F$.

Let F_0 denote the fixed field of F. There are two important maps from F to F_0 associated with σ, the *trace* and *norm* maps, defined by

$$\begin{aligned} \mathrm{Tr}(c) &= c + \overline{c}, \\ N(c) &= c \cdot \overline{c}. \end{aligned}$$

Now Tr is an additive homomorphism (indeed, an F_0-linear map), and N is a multiplicative homomorphism. As we have seen, the image of Tr is F_0; the kernel is the set of c such that $c^\sigma = -c$ (which is equal to F_0 if the characteristic is 2 but not otherwise).

Suppose that F is finite. Then the order of F is a square, say $F = \mathrm{GF}(q^2)$, and $F_0 = \mathrm{GF}(q)$. Since the multiplicative group of F_0 has order $q-1$, a non-zero element $c \in F$ lies in F_0 if and only if $c^{q-1} = 1$. This holds if and only if $c = a^{q+1}$ for some $a \in F$ (as the multiplicative group of F is cyclic), in other words, $c = a \cdot \overline{a} = N(a)$. So the image of N is the multiplicative group of F_0, and its kernel is the set of $(q+1)$st roots of 1. Also, the kernel of Tr consists of zero and the set of $(q-1)$st roots of -1, the latter being a coset of $F_0^\times$ in $F^\times$.

The Hermitian form on a hyperbolic plane has the form

$$B(x,y) = x_1\overline{y_2} + y_1\overline{x_2}.$$

An arbitrary Hermitian formed space is the orthogonal direct sum of r hyperbolic planes and an anisotropic space. We have seen that, up to scalar multiplication, the following hold:

(a) over $\mathbb{C}$, an anisotropic space is positive definite, and the form can be taken to be

$$B(x,y) = x_1\overline{y_1} + \cdots + x_s\overline{y_s};$$

(b) over a finite field, an anisotropic space has dimension at most one; if non-zero, the form can be taken to be

$$B(x,y) = x\overline{y}.$$

5.1 The unitary groups

Let A be the matrix associated with a non-degenerate Hermitian form B. Then $A = \overline{A}^\top$, and the isometry group of B (the *unitary group* $\mathrm{U}(V,B)$ consists of all invertible matrices P which satisfy $\overline{P}^\top AP = A$.

Since A is invertible, we see that

$$N(\det(P)) = \det(\overline{P}^\top)\det(P) = 1.$$

So $\det(P) \in F_0$. Moreover, a scalar matrix cI lies in the unitary group if and only if $N(c) = c\overline{c} = 1$.

The *special unitary group* $\mathrm{SU}(V,B)$ consists of all elements of the unitary group which have determinant 1 (that is, $\mathrm{SU}(V,B) = \mathrm{U}(V,B) \cap \mathrm{SL}(V)$), and the *projective special unitary group* is the factor group $\mathrm{SU}(V,B)/SU(V,B) \cap Z$, where Z is the group of scalar matrices.

In the case where $F = \mathrm{GF}(q^2)$ is finite, we can unambiguously write $\mathrm{SU}(n,q)$ and $\mathrm{PSU}(n,q)$, since up to scalar multiplication there is a unique Hermitian form on $\mathrm{GF}(q^2)^n$ (with rank $\lfloor n/2 \rfloor$ and germ of dimension 0 or 1 according as n is even or odd). (It would be more logical to write $\mathrm{SU}(n,q^2)$ and $\mathrm{PSU}(n,q^2)$ for these groups; we have used the standard group-theoretic convention.

Proposition 5.1 *(a)* $|\,\mathrm{U}(n,q) = q^{n(n-1)/2}\prod_{i=1}^{n}(q^i - (-1)^i)$.

(b) $|\,\mathrm{SU}(n,q)| = |\,\mathrm{U}(n,q)|/(q+1)$.

(c) $|\,\mathrm{PSU}(n,q)| = |\,\mathrm{SU}(n,q)|/d$*, where* $d = (n,q+1)$.

Proof (a) We use Theorem 3.17, with either $n = 2r$, $\varepsilon = -\frac{1}{2}$, or $n = 2r+1$, $\varepsilon = \frac{1}{2}$, and with q replaced by q^2, noting that, in the latter case, $|G_0| = q+1$. It happens that both cases can be expressed by the same formula! On the same theme, note that, if we replace $(-1)^i$ by 1 (and $q+1$ by $q-1$ in parts (b) and (c) of the theorem), we obtain the orders of $\mathrm{GL}(n,q)$, $\mathrm{SL}(n,q)$, and $\mathrm{PSL}(n,q)$ instead.

(b) As we noted, det is a homomorphism from $\mathrm{U}(n,q)$ onto the group of $(q+1)$st roots of unity in $\mathrm{GF}(q^2)^\times$, whose kernel is $\mathrm{SU}(n,q)$.

(c) A scalar cI belongs to $\mathrm{U}(n,q)$ if $c^{q+1} = 1$, and to $\mathrm{SL}(n,q^2)$ if $c^n = 1$. So $|Z \cap \mathrm{SL}(n,q^2)| = d$, as required.

We conclude this section by considering unitary transvections, those which preserve a Hermitian form. Accordingly, let $T : x \mapsto x + (xf)a$ be a transvection, where $af = 0$. We have

$$\begin{array}{rcl} B(xT, yT) & = & B(x + (xf)a, y + (yf)a) \\ & = & B(x,y) + (xf)\overline{B(y,a)} + \overline{(yf)}B(x,a) + (xf)\overline{(yf)}B(a,a). \end{array}$$

So T is unitary if and only if the last three terms vanish for all x, y. Putting $y = a$ we see that $(xf)\overline{B(a,a)} = 0$ for all x, whence (since $f \neq 0$) we must have $B(a,a) = 0$. Now choosing y such that $B(y,a) = 1$ and setting $\lambda = \overline{(yf)}$, we have $xf = \lambda B(x,a)$ for all x. So a unitary transvection has the form

$$x \mapsto x + \lambda B(x,a)a,$$

where $B(a,a) = 0$. In particular, an anisotropic space admits no unitary transvections. Also, choosing x and y such that $B(x,a) = B(y,a) = 1$, we find that $\mathrm{Tr}(\lambda) = 0$. Conversely, for any $\lambda \in \ker(\mathrm{Tr})$ and any a with $B(a,a) = 0$, the above formula defines a unitary transvection.

5.2 Hyperbolic planes

In this section only, we use the convention that $\mathrm{U}(2, F_0)$ means the unitary group associated with a hyperbolic plane over F, and σ is the associated field automorphism, having fixed field F_0.

Theorem 5.2 $\mathrm{SU}(2, F_0) \cong \mathrm{SL}(2, F_0)$.

Proof We will show, moreover, that the actions of the unitary group on the polar space and that of the special linear group on the projective space correspond, and that unitary transvections correspond to transvections in $\mathrm{SL}(2, F_0)$. Let $K = \{c \in F : c + \overline{c} = 0\}$ be the kernel of the trace map; recall that the image of the trace map is F_0.

With the standard hyperbolic form, we find that a unitary matrix

$$P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

must satisfy $\overline{P}^{\top} AP = A$, where

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Hence

$$a\overline{c}+\overline{a}c=0, \quad b\overline{c}+\overline{a}d=1, \quad b\overline{d}+\overline{b}d=0.$$

In addition, we require, that $\det(P)=1$, that is, $ad-bc=1$.

From these equations we deduce that $b+\overline{b}=c+\overline{c}=0$, that is, $b,c\in K$, while $a-\overline{a}=d-\overline{d}=0$, that is, $a,d\in F_0$.

Choose a fixed element $u\in K$. Then $\lambda\in K$ if and only if $u\lambda\in F_0$. Also, $u^{-1}\in K$. Hence the matrix

$$P^\dagger=\begin{pmatrix} a & ub \\ u^{-1}c & d\end{pmatrix}$$

belongs to $\mathrm{SL}(2,F_0)$. Conversely, any matrix in $\mathrm{SL}(2,F_0)$ gives rise to a matrix in $\mathrm{SU}(2,F_0)$ by the inverse map. So we have a bijection between the two groups. It is now routine to check that the map is an isomorphism.

Represent the points of the projective line over F by $F\cup\{\infty\}$ as usual. Recall that ∞ is the point (rank 1 subspace) spanned by $(0,1)$, while c is the point spanned by $(1,c)$. We see that ∞ is flat, while c is flat if and only if $c+\overline{c}=0$, that is, $c\in K$. So the map $x\mapsto x$ takes the polar space for the unitary group onto the projective line over F_0. It is readily checked that this map takes the action of the unitary group to that of the special linear group.

By transitivity, it is enough to consider the unitary transvections $x\mapsto x+\lambda B(x,a)a$, where $a=(0,1)$. In matrix form, these are

$$P=\begin{pmatrix} 1 & \lambda \\ 0 & 1\end{pmatrix},$$

with $\lambda\in K$. Then

$$P^\dagger=\begin{pmatrix} 1 & u\lambda \\ 0 & 1\end{pmatrix},$$

which is a transvection in $\mathrm{SL}(2,F_0)$, as required. ∎

In particular, we see that $\mathrm{PSU}(2,F_0)$ is simple if $|F_0|>3$.

5.3 Generation and simplicity

We follow the now-familiar pattern. First we treat two exceptional finite groups, then we show that unitary groups are generated by unitary transvections and that most are simple. By the preceding section, we may assume that the rank is at least 3.

The finite unitary group $\mathrm{PSU}(3,q)$ is a 2-transitive group of permutations of the q^3+1 points of the corresponding polar space (since any two such points are spanned by a hyperbolic pair) and has order $(q^3+1)q^3(q^2-1)/d$, where $d = (3,q+1)$. Moreover, any two points span a line containing $q+1$ points of the polar space. The corresponding geometry is called a *unital*.

For $q = 2$, the group has order 72, and so is soluble. In fact, it is *sharply 2-transitive*: a unique group element carries any pair of points to any other.

Exercise 5.1 (a) Show that the unital associated with $\mathrm{PSU}(3,2)$ is isomorphic to the *affine plane* over $\mathrm{GF}(3)$, defined as follows: the points are the vectors in a vector space V of rank 2 over $\mathrm{GF}(3)$, and the lines are the cosets of rank 1 subspaces of V (which, over the field $\mathrm{GF}(3)$, means the triples of vectors with sum 0).

(b) Show that the automorphism group of the unital has the structure $3^2:\mathrm{GL}(2,3)$, where 3^2 denotes an elementary abelian group of this order (the translation group of V) and : denotes semidirect product.

(c) Show that $\mathrm{PSU}(3,2)$ is isomorphic to $3^2 : Q_8$, where Q_8 is the quaternion group of order 8.

(d) Show that $\mathrm{PSU}(3,2)$ is not generated by unitary transvections.

We next consider the group $\mathrm{PSU}(4,2)$, and outline the proof of the following theorem:

Theorem 5.3 $\mathrm{PSU}(4,2) \cong \mathrm{PSp}(4,3)$.

Proof Observe first that both these groups have order 25 920. We will construct a geometry for the group $\mathrm{PSU}(4,2)$, and use the technical results of Section 4.4 to identify it with the generalised quadrangle for $\mathrm{PSp}(4,3)$. Now it has index 2 in the full automorphism group of this geometry, as also does $\mathrm{PSp}(4,3)$, which is simple; so these two groups must coincide.

The geometry is constructed as follows. Let V be a vector space of rank 4 over $\mathrm{GF}(4)$ carrying a Hermitian form of polar rank 2. The projective space $\mathrm{PG}(3,4)$ derived from V has $(4^4-1)/(4-1) = 85$ points, of which $(4^2-1)(4^{3/2}+1)/(4-1) = 45$ are points of the polar space, and the remaining 40 are points on which the form does not vanish (spanned by vectors x with $B(x,x) = 1$). Note that $40 = (3^4-1)/(3-1)$ is equal to the number of points of the symplectic generalised quadrangle over $\mathrm{GF}(3)$. Let Ω denote this set of 40 points.

Define an F-line to be a set of four points of Ω spanned by the vectors of an orthonormal basis for V (a set of four vectors x_1, x_2, x_3, x_4 with $B(x_i, x_i) = 1$ and $B(x_i, x_j) = 0$ for $i \neq j$). Note that two orthogonal points p, q of Ω span a non-degenerate 2-space, which is a line containing five points of the projective space of which three are flat and the other two belong to Ω. Then $\{p, q\}^{\perp}$ is also a non-degenerate 2-space containing two points of Ω, which complete $\{p, q\}$ to an F-line. Thus, two orthogonal points lie on a unique F-line, while two non-orthogonal points lie on no F-line. It is readily checked that, if $L = \{p_1, p_2, p_3, p_4\}$ is an F-line and q is another point of Ω, then p has three non-zero coordinates in the orthonormal basis corresponding to L, so q is orthogonal to a unique point of L. Thus, the points of Ω and the F-lines satisfy condition (a) of Section 4.4; that is, they form a generalised quadrangle.

Now consider two points of Ω which are not orthogonal. The 2-space they span is degenerate, with a radical of rank 1. So of the five points of the corresponding projective line, four lie in Ω and one (the radical) is flat. Sets of four points of this type (which are obviously determined by any two of their members) will be the H-lines. It is readily checked that the H-lines do indeed arise in the manner described in Section 4.4, that is, as the sets of points of Ω orthogonal to two given non-orthogonal points. So condition (b) holds.

Now a point p of Ω lies in four F-lines, whose union consists of thirteen points. If q and r are two of these points which do not lie on an F-line with p, then q and r cannot be orthogonal, and so they lie in an H-line; since p and q are orthogonal to p, so are the remaining points of the H-line containing them. Thus we have condition (c). Now (d) is easily verified by counting, and the proof is complete. ∎

Exercise 5.2 (a) Give a detailed proof of the above isomorphism.

(b) If you are familiar with a computer algebra package, verify computationally that the above geometry for $\mathrm{PSU}(4,2)$ is isomorphic to the symplectic generalised quadrangle for $\mathrm{PSp}(4,3)$.

In our generation and simplicity results we treat the rank 3 case separately. In the rank 3 case, the unitary group is 2-transitive on the points of the unital.

Theorem 5.4 *Let (V, B) be a unitary formed space of Witt rank 1, with $\mathrm{rk}(V) = 3$. Assume that the field F is not $\mathrm{GF}(2^2)$.*

(a) $\mathrm{SU}(V, B)$ *is generated by unitary transvections.*

(b) $\mathrm{PSU}(V, B)$ *is simple.*

Proof We exclude the case of $\mathrm{PSU}(3,2)$ (with $F = \mathrm{GF}(2^2)$, considered earlier. Replacing the form by a scalar multiple if necessary, we assume that the germ contains vectors of norm 1. Take such a vector as second basis vector, where the first and third are a hyperbolic pair. That is, we assume that the form is

$$B((x_1,x_2,x_3),(y_1,y_2,y_3)) = x_1\overline{y_3}+x_2\overline{y_2}+x_3\overline{y_1},$$

so the isometry group is

$$\{P : \overline{P}^\top AP = A\}$$

where

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Now we check that the group

$$Q = \left\{ \begin{pmatrix} 1 & -\overline{a} & b \\ 0 & 1 & a \\ 0 & 0 & 1 \end{pmatrix} : N(a)+\mathrm{Tr}(b) = 0 \right\}$$

is a subgroup of $G = \mathrm{SU}(V,B)$, and its derived group consists of unitary transvections (the elements with $a = 0$).

Next we show that the subgroup T of V generated by the transvections in G is transitive on the set of vectors x such that $B(x,x) = 1$. Let x and y be two such vectors. Suppose first that $\langle x,y\rangle$ is nondegenerate. Then it is a hyperbolic line, and a calculation in $\mathrm{SU}(2,F_0)$ gives the result. Otherwise, there exists z such that $\langle x,z\rangle$ and $\langle y,z\rangle$ are nondegenerate, so we can get from x to y in two steps.

Now the stabiliser of such a vector in G is $\mathrm{SU}(x^\perp,B) = \mathrm{SU}(2,F_0)$, which is generated by transvections; and every coset of this stabiliser contains a transvection. So G is generated by transvections.

Now it follows that the transvections lie in G', and Iwasawa's Lemma (Theorem 2.7) shows that $\mathrm{PSL}(V,B)$ is simple. ■

Exercise 5.3 Complete the details in the above proof by showing

(a) the group $\mathrm{SU}(2,F_0)$ acts transitively on the set of vectors of norm 1 in the hyperbolic plane;

(b) given two vectors x,y of norm 1 in a rank 3 unitary space as in the proof, either $\langle x,y\rangle$ is a hyperbolic plane, or there exists z such that $\langle x,z\rangle$ and $\langle y,z\rangle$ are hyperbolic planes.

Theorem 5.5 *Let (V,B) be a unitary formed space with Witt rank at least 2. Then*

(a) $\mathrm{SU}(V,B)$ *is generated by unitary transvections.*

(b) $\mathrm{PSU}(V,B)$ *is simple.*

Proof We follow the usual pattern. The argument in the preceding theorem shows part (a) without change if $F \neq \mathrm{GF}(4)$. In the excluded case, we know that $\mathrm{PSU}(4,2) \cong \mathrm{PSp}(4,3)$ is simple, and so is generated by any conjugacy class (in particular, the images of the transvections of $\mathrm{SU}(4,2)$). Then induction shows the result for higher rank spaces over $\mathrm{GF}(4)$. Again, the argument in 3 dimensions shows that transvections are commutators; the action on the points of the polar space is primitive; and so Iwasawa's Lemma shows the simplicity. ■

6 Orthogonal groups

We now turn to the orthogonal groups. These are more difficult, for two related reasons. First, it is not always true that the group of isometries with determinant 1 is equal to its derived group (and simple modulo scalars). Secondly, in characteristic different from 2, there are no transvections, and we have to use a different class of elements.

We let $\mathrm{O}(Q)$ denote the isometry group of the non-degenerate quadratic form Q, and $\mathrm{SO}(Q)$the group of isometries with determinant 1. Further, $\mathrm{PO}(Q)$ and $\mathrm{PSO}(Q)$ are the quotients of these groups by the scalars they contain. We define $\Omega(Q)$ to be the derived subgroup of $\mathrm{O}(Q)$, and $\mathrm{P}\Omega(Q) = \Omega(Q)/(\Omega(Q) \cap Z)$, where Z consists of the scalar matrices. Sometimes $\Omega(Q) = \mathrm{SO}(Q)$, and sometimes it is strictly smaller; but our notation serves for both cases.

In the case where F is finite, we have seen that for even n there is a unique type of non-degenerate quadratic form up to scalar multiplication, while if n is even there are two types, having germ of dimension 0 or 2 respectively, We write $\mathrm{O}^+(n,q)$, $\mathrm{O}(n,q)$ and $\mathrm{O}^-(n,q)$ for the isometry group of a non-degenerate quadratic form on $\mathrm{GF}(q)^n$ with germ of rank 0, 1, 2 (and n even, odd, even respectively). We use similar notation for SO, PΩ, and so on. Then we write $\mathrm{O}^\varepsilon(n,q)$ to mean either $\mathrm{O}^+(n,q)$ or $\mathrm{O}^-(n,q)$. Note that, unfortunately, this convention (which is standard notation) makes ε the negative of the ε appearing in our general order formula (Theorem 3.17).

Now the order formula for the finite orthogonal groups reads as follows.

$$
\begin{aligned}
|\,\mathrm{O}(2m+1,q)| &= d\prod_{i=1}^{m}(q^{2i}-1)q^{2i-1} \\
&= dq^{m^2}\prod_{i=1}^{m}(q^{2i}-1), \\
|\,\mathrm{O}^+(2m,q)| &= \prod_{i=1}^{m}(q^{i}-1)(q^{i-1}+1)q^{2i-2} \\
&= 2q^{m(m-1)}(q^m-1)\prod_{i=1}^{m-1}(q^{2i}-1), \\
|\,\mathrm{O}^-(2m,q)| &= 2(q+1)\prod_{i=1}^{m-1}(q^{i}-1)(q^{i+1}+1)q^{2i}
\end{aligned}
$$

$$= \quad 2q^{m(m-1)}(q^m+1)\prod_{i=1}^{m-1}(q^{2i}-1),$$

where $d=(2,q-1)$. Note that there is a single difference in sign between the final formulae for $\mathrm{O}^{\varepsilon}(2m,q)$ for $\varepsilon=\pm 1$; we can combine the two and write

$$|\,\mathrm{O}^{\varepsilon}(2m,q)\,| = 2(q^m-\varepsilon)\prod_{i=1}^{m-1}(q^{2i}-1).$$

We have $|\,\mathrm{SO}(n,q)\,| = |\,\mathrm{O}(n,q)\,|/d$ (except possibly if n is odd and q is even). This is because, with this exclusion, the bilinear form B associated with Q is non-degenerate; and the orthogonal group consists of matrices P satisfying $P^{\top}AP=A$, where A is the matrix of the bilinear form, so that $\det(P)=\pm 1$. It is easy to show that, for q odd, there are orthogonal transformations with determinant -1. The excluded case will be dealt with in Section 6.2. We see also that the only scalars in $\mathrm{O}(Q)$ are $\pm I$; and, in characteristic different from 2, we have $-I\in\mathrm{SO}(Q)$ if and only if the rank of the underlying vector space is even. Thus, for q odd, we have

$$|\,\mathrm{SO}(Q)\,| = |\,\mathrm{PO}(Q)\,| = |\,\mathrm{O}(Q)\,|/2,$$

and

$$|\,\mathrm{PSO}(Q)\,| = |\,\mathrm{SO}(Q)\,|/(n,2).$$

For q and n even we have $\mathrm{O}(Q)=\mathrm{SO}(Q)=\mathrm{PO}(Q)=\mathrm{PSO}(Q)$.

Exercise 6.1 Let Q be a non-degenerate quadratic form over a field of characteristic different from 2. Show that $\mathrm{O}(Q)$ contains elements with determinant -1. [Hint: if $Q(v)\neq 0$, then take the transformation which takes v to $-v$ and extend it by the identity on $v^{\perp}$ (in other words, the reflection in the hyperplane $v^{\perp}$).]

6.1 Some small-dimensional cases

We begin by considering some small cases. Let V be a vector space of rank n carrying a quadratic form Q of Witt index r, where $\delta=n-2r$ is the dimension of the germ of Q. Let $\mathrm{O}(Q)$ denote the isometry group of Q, and $\mathrm{SO}(Q)$ the subgroup of isometries of determinant 1.

Case $n=1$, $r=0$. In this case the quadratic form is a scalar multiple of x^2. (Replacing q by a scalar multiple does not affect the isometry group.) Then $\mathrm{O}(Q)=\{\pm 1\}$, a group of order 1 or 2 according as the characteristic is or is not 2; and $\mathrm{SO}(Q)$ is the trivial group.

Case $n = 2$, $r = 1$. The quadratic form is $Q(x_1, x_2) = x_1x_2$, and the isometry group $G = \mathrm{O}(Q)$ is

$$\left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \begin{pmatrix} 0 & \lambda \\ \lambda^{-1} & 0 \end{pmatrix} : \lambda \in F^\times \right\},$$

a group with a subgroup H of index 2 isomorphic to $F^\times$, and such that an element $t \in G \setminus H$ satisfies $t^2 = 1$ and $t^{-1}ht = h^{-1}$ for all $h \in H$. In other words, G is a *generalised dihedral group*. If $F = \mathrm{GF}(q)$, then $\mathrm{O}(2,q)$ is a dihedral group of order $2(q-1)$. Note that $H = \mathrm{SO}(Q)$ if and only if the characteristic of F is not 2.

Case $n = 2$, $r = 0$. In this case, the quadratic form is

$$\alpha x_1^2 + \beta x_1 x_2 + \gamma x_2^2,$$

where $q(x) = \alpha x^2 + \beta x + \gamma$ is an irreducible quadratic over F. Let K be a splitting field for q over F, and *assume* that K is a Galois extension (in other words, that q is separable over F: this includes the cases where either the characteristic is not 2 or the field F is finite). Then, up to scalar multiplication, the form Q is equivalent to the K/F norm on the F-vector space K. The orthogonal group is generated by the multiplicative group of elements of norm 1 in K and the Galois automorphism σ of K over F.

In the case $F = \mathrm{GF}(q)$, this group is dihedral of order $2(q+1)$. In the case $F = \mathbb{R}$, the $\mathbb{C}/\mathbb{R}$ norm is

$$z \mapsto z\bar{z} = |z|^2,$$

and so the orthogonal group is generated by multiplication by unit complex numbers and complex conjugation. In geometric terms, it is the group of rotations and reflections about the origin in the Euclidean plane.

Again we see that $\mathrm{SO}(Q)$ has index 2 in $\mathrm{O}(F)$ if the characteristic of F is not 2.

Exercise 6.2 Prove that, if K is a Galois extension of F, then the determinant of the F-linear map $x \mapsto \lambda x$ on K is equal to $N_{K/F}(\lambda)$. [Hint: if $\lambda \notin F$, the eigenvalues of this map are λ and λ^σ.]

Case $n = 3$, $r = 1$. In this case and the next, we describe a group preserving a quadratic form and claim without proof that it is the full orthogonal group. Also, in this case, we assume that the characteristic is not equal to 2.

Let $V = F^2$, and let W be the vector space of all quadratic forms on V (not necessarily non-degenerate). Then $\mathrm{rk}(W) = 3$; a typical element of W is the quadratic form $ux^2 + vxy + wy^2$, where we have represented a typical vector in V as (x, y). We use the triple (u, v, w) of coefficients to represent this vector of w. Now $\mathrm{GL}(V)$ acts on W by substitution on the variables in the quadratic form. In other words, to the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(2, F)$$

corresponds the map

$$\begin{aligned} ux^2 + vxy + wy^2 &\mapsto u(ax + cy)^2 + v(ax + cy)(bx + dy) + w(bx + dy)^2 \\ &= (ua^2 + vab + wb^2)x^2 + (2uac + v(ad + bc) + 2wbd)xy \\ &\qquad + (uc^2 + vcd + wd^2)y^2, \end{aligned}$$

which is represented by the matrix

$$\rho(A) = \begin{pmatrix} a^2 & 2ac & c^2 \\ ab & ad + bc & cd \\ b^2 & 2bd & d^2 \end{pmatrix} \in \mathrm{GL}(3, F).$$

We observe several things about this representation ρ of $\mathrm{GL}(2, F)$:

(a) The kernel of the representation is $\{\pm I\}$.

(b) $\det(\rho(A)) = (\det(A))^3$.

(c) The quadratic form $Q(u, v, w) = 4uw - v^2$ is multiplied by a factor $\det(A)^2$ by the action of $\rho(A)$.

Hence we find a subgroup of $\mathrm{O}(Q)$ which is isomorphic to $\mathrm{SL}^{\pm}(2, F)/\{\pm I\}$, where $\mathrm{SL}^{\pm}(2, F)$ is the group of matrices with determinant ± 1. Moreover, its intersection with $\mathrm{SL}(3, F)$ is $\mathrm{SL}(2, F)/\{\pm 1\}$. In fact, these are the full groups $\mathrm{O}(Q)$ and $\mathrm{SO}(Q)$ respectively.

We see that in this case,

$$\mathrm{P}\Omega(Q) \cong \mathrm{PSL}(2, F) \quad \text{if } |F| > 2,$$

and this group is simple if $|F| > 3$.

Case $n = 4$, $r = 2$. Our strategy is similar. We take the rank 4 vector space over F to be the space $M^{2\times 2}(F)$, the space of 2×2 matrices over F (where F is any field). The determinant function on V is a quadratic form: $Q(X) = \det(X)$. Clearly X is the sum of two hyperbolic planes (for example, the diagonal and the antidiagonal matrices).

There is an action of the group $\mathrm{GL}(2,F)\times\mathrm{GL}(2,F)$ on X, by the rule

$$\rho(A,B) : X \mapsto A^{-1}XB.$$

We see that $\rho(A,B)$ preserves Q if and only if $\det(A) = \det(B)$, and $\rho(A,B)$ is the identity if and only if $A = B = \lambda I$ for some scalar λ. So we have a subgroup of $\mathrm{O}(Q)$ with the structure

$$((\mathrm{SL}(2,F)\times\mathrm{SL}(2,F))\cdot F^{\times})/\{(\lambda I,\lambda I) : \lambda\in F^{\times}\}.$$

Moreover, the map $T : X \mapsto X^{\top}$ also preserves Q. It can be shown that together these elements generate $\mathrm{O}(Q)$.

Exercise 6.3 Show that the above map T has determinant -1 on V, while $\rho(A,B)$ has determinant equal to $\det(A)^{-2}\det(B)^2$. Deduce (from the information given) that $\mathrm{SO}(Q)$ has index 2 in $O(Q)$ if and only if the characteristic of F is not 2.

Exercise 6.4 Show that, in the above case, we have

$$\mathrm{P\Omega}(Q) \cong \mathrm{PSL}(2,F)\times\mathrm{PSL}(2,F)$$

if $|F| > 3$.

Exercise 6.5 Use the order formulae for finite orthogonal groups to prove that the groups constructed on vector spaces of ranks 3 and 4 are the full orthogonal groups, as claimed.

6.2 Characteristic 2, odd rank

In the case where the bilinear form is degenerate, we show that the orthogonal group is isomorphic to a symplectic group.

Theorem 6.1 *Let F be a perfect field of characteristic 2. Let Q be a non-degenerate quadratic form in n variables over F, where n is odd. Then* $\mathrm{O}(Q)\cong\mathrm{Sp}(n-1,F)$.

Proof We know that the bilinear form B is alternating and has a rank 1 radical, spanned by a vector z, say. By multiplying Q by a scalar if necessary, we may assume that $Q(z) = 1$. Let $\overline{G}$ be the group induced on V/Z, where $Z = \langle z \rangle$. Then $\overline{G}$ preserves the symplectic form.

The kernel K of the homomorphism from G to $\overline{G}$ fixes each coset of Z. Since

$$Q(v + az) = Q(v) + a^2,$$

and the map $a \mapsto a^2$ is a bijection of F, each coset of Z contains one vector with each possible value of Q. Thus $K = 1$, and $G \cong \overline{G}$.

Conversely, let $\overline{g}$ be any linear transformation of V/Z which preserves the symplectic form induced by B. The above argument shows that there is a unique permutation g of V lifting the action of $\overline{g}$ and preserving Q. Note that, since $\overline{g}$ induces g on V/Z, it preserves B. We claim that g is linear. First, take any two vectors v, w. Then

$$\begin{aligned} Q(vg + wg) &= Q(vg) + Q(wg) + B(vg, wg) \\ &= Q(v) + Q(w) + B(v, w) \\ &= Q(v + w) \\ &= Q((v + w)g); \end{aligned}$$

and the linearity of $\overline{g}$ shows that $vg + wg$ and $(v + w)g$ belong to the same coset of Z, and so they are equal. A similar argument applies for scalar multiplication. So $\overline{G} = \mathrm{Sp}(n-1, F)$, and the result is proved. ■

We conclude that, with the hypotheses of the theorem, $\mathrm{O}(Q)$ is simple except for $n = 3$ or $n = 5$, $F = \mathrm{GF}(2)$. Hence $\mathrm{O}(Q)$ coincides with $\mathrm{P}\Omega(Q)$ with these exceptions.

We conclude by constructing some more 2-transitive groups. Let F be a perfect field of characteristic 2, and B a symplectic form on F^{2m}. Then the set $\mathcal{Q}(B)$ of all quadratic forms which polarise to B is a coset of the set of "square-seminilear maps" on V, those satisfying

$$\begin{aligned} L(x + y) &= L(x) + L(y), \\ L(cx) &= c^2 L(x) \end{aligned}$$

(these maps are just the quadratic forms which polarise to the zero bilinear form).

In the finite case, where $F = \mathrm{GF}(q)$ (q even), there are thus q^{2m} such quadratic forms, and they fall into two orbits under $\mathrm{Sp}(2m, q)$, corresponding to the two types of forms. The stabiliser of a form Q is the corresponding orthogonal group $\mathrm{O}(Q)$. The number of forms of each type is the index of the corresponding orthogonal group in the symplectic group, which can be calculated to be $q^m(q^m+\varepsilon)/2$ for a form of type ε.

Now specialise further to $F = \mathrm{GF}(2)$. In this case, "square-semilinear" maps are linear. So, given a quadratic form Q polarising to B, we have

$$\mathcal{Q}(B) = \{Q + L : L \in V^*\}.$$

Further, each linear form can be written as $x \mapsto B(x,a)$ for some fixed $a \in V$. Thus, there is an $\mathrm{O}(Q)$-invariant bijection between $\mathcal{Q}(B)$ and V. By Witt's Lemma, $\mathrm{O}(Q)$ has just three orbits on V, namely

$$\{0\}, \quad \{x \in V : Q(x) = 0, x \neq 0\}, \quad \{x \in V : Q(x) = 1\}.$$

So $\mathrm{O}(Q)$ has just three orbits on $\mathcal{Q}(B)$, namely

$$\{Q\}, \quad \mathcal{Q}^{\varepsilon}(B) \setminus \{Q\}, \quad \mathcal{Q}^{-\varepsilon}(B),$$

where Q has type ε and $\mathcal{Q}^{\varepsilon}(B)$ is the set of all forms of type ε in $\mathcal{Q}(B)$.

It follows that $\mathrm{Sp}(2m,2)$ acts 2-transitively on each of the two sets $\mathcal{Q}^{\varepsilon}(B)$, with cardinalities $2^{m-1}(2^m+\varepsilon)$. The point stabiliser in these actions are $\mathrm{O}^{\varepsilon}(2m,2)$.

Exercise 6.6 What isomorphisms between symmetric and classical groups are illustrated by the above 2-transitive actions of $\mathrm{Sp}(4,2)$?

6.3 Transvections and root elements

We first investigate *orthogonal transvections*, those which preserve the non-degenerate quadratic form Q on the F-vector space V.

Proposition 6.2 *There are no orthogonal transvections over a field F whose characteristic is different from 2. If F has characteristic 2, then an orthogonal transvection for a quadratic form Q has the form*

$$x \mapsto x - Q(a)^{-1}B(x,a)a,$$

where $Q(a) \neq 0$ and B is obtained by polarising Q.

Proof Suppose that the transvection $x \mapsto x + (xf)a$ preserves the quadratic form Q, and let B be the associated bilinear form. Then

$$Q(x + (xf)a) = Q(x)$$

for all $x \in V$, whence

$$(xf)^2 Q(a) + (xf)B(x,a) = 0.$$

If $xf \neq 0$, we conclude that $(xf)Q(a) + B(x,a) = 0$. Since this linear equation holds on the complement of a hyperplane, it holds everywhere; that is, $B(x,a) = -(xf)Q(a)$ for all x.

If the characteristic is not 2, then $B(a,a) = 2Q(a)$. Substituting $x = a$ in the above equation, using $af = 0$, we see that $B(a,a) = 0$, so $Q(a) = 0$. But then $B(x,a) = 0$ for all x, contradicting the nondegeneracy of B in this case.

So we may assume that the characteristic is 2. If $B(x,a) = 0$ for all $x \in V$, then $Q(a) \neq 0$ by non-degeneracy. Otherwise, choosing x with $B(x,a) \neq 0$, we see that again $Q(a) \neq 0$. Then

$$xf = -Q(a)^{-1} B(x,a),$$

and the proof is complete. (Incidentally, the fact that f is non-zero now shows that a is not in the radical of B.)

Exercise 6.7 In the characteristic 2 case, replacing a by λa for $a \neq 0$ does not change the orthogonal transvection.

The fact that, if Q is a non-degenerate quadratic form in three variables with Witt rank 1 shows that we can find analogues of transvections acting on three-dimensional sections of V. These are called *root elements*, and they will be used in our simplicity proofs.

A *root element* is a transformation of the form

$$x \mapsto x + aB(x,v)u - aB(x,u)v - a^2 Q(v)B(x,u)u$$

where $Q(u) = B(u,v) = 0$. The group of all such transformations for fixed u, v satisfying the above conditions, together with the identity, is called a *root subgroup* $X_{u,v}$.

Exercise 6.8 Prove that the root elements are isometries of Q, that they have determinant 1, and that the root subgroups are abelian. Show further that, if $Q(u) = 0$, then the group

$$X_u = \langle X_{u,v} : v \in u^{\perp} \rangle$$

is abelian, and is isomorphic to the additive group of $u^{\perp}/\langle u \rangle$.

Exercise 6.9 Write down the root subgroup $X_{u,v}$ for the quadratic form $Q(x_1,x_2,x_3) = x_1x_3 - x_2^2$ relative to to the given basis $\{e_1,e_2,e_3\}$, where $u = e_1$ and $v = e_2$.

Now the details needed to apply Iwasawa's Lemma are similar to, but more complicated than, those that we have seen in the cases of the other classical groups. We summarise the important steps. Let Q be a quadratic form with Witt rank at least 2, and not of Witt rank 2 on a vector space of rank 4 (that is, not equivalent to $x_1x_2 + x_3x_4$). We also exclude the case where Q has Witt index 2 on a rank 5 vector space over $\mathrm{GF}(2)$: in this case $\mathrm{P}\Omega(Q) \cong \mathrm{PSp}(4,2) \cong S_6$.

(a) The root subgroups are contained in $\Omega(Q)$, the derived group of $O(Q)$.

(b) The abelian group X_u is normal in the stabiliser of u.

(c) $\Omega(Q)$ is generated by the root subgroups.

(d) $\Omega(Q)$ acts primitively on the set of flat 1-spaces.

Note that the exception of the case of rank 4 and Witt index 2 is really necessary for (d): the group $\Omega(Q)$ fixes the two families of rulings on the hyperbolic quadric shown in Figure 1 on p. 41, and each family is a system of blocks of imprimitivity for this group.

Then from Iwasawa's Lemma we conclude:

Theorem 6.3 *Let Q be a non-degenerate quadratic form with Witt rank at least 2, but not of Witt rank 2 on either a vector space of rank 4 or a vector space of rank 5 over $\mathrm{GF}(2)$. Then $\mathrm{P}\Omega(Q)$ is simple.*

It remains for us to discover the order of $\mathrm{P}\Omega(Q)$ over a finite field. We give the result here, and defer the proof until later. The facts are as follows.

Proposition 6.4 *(a) Let Q have Witt index at least 2, and let F have characteristic different from 2. Then $\mathrm{SO}(Q)/\Omega(Q) \cong F^\times/(F^\times)^2$.*

(b) Let F be a perfect field of characteristic 2 and let Q have Witt index at least 2; exclude the case of a rank 4 vector space over $\mathrm{GF}(2)$. Then $\mathrm{SO}(Q) : \Omega(Q)| = 2$.

The proof of part (a) involves defining a homomorphism from $\mathrm{SO}(Q)$ to $F^\times/(F^\times)^2$ called the *spinor norm*, and showing that it is onto and its kernel is $\Omega(Q)$ except in the excluded case.

In the remaining cases, we work over the finite field $\mathrm{GF}(q)$, and write $O(n,q)$, understanding that if n is even then $O^\varepsilon(n,q)$ is meant.

Proposition 6.5 *Excluding the case q even and n odd:*

(a) $|\,\mathrm{SO}(n,q):\Omega(n,q)| = 2$.

(b) For q odd, $-I \in \Omega^\varepsilon(2m,q)$ *if and only if* $q^m \cong \varepsilon \pmod 4$.

The last part is proved by calculating the spinor norm of $-I$. Putting this together with the order formula for $\mathrm{SO}(n,q)$ already noted, we obtain the following result:

Theorem 6.6 *For* $m \geq 2$*, excluding the case* $\mathrm{P}\Omega^+(4,2)$*, we have*

$$\begin{aligned}
|\,\mathrm{P}\Omega^\varepsilon(2m,q)| &= \left(q^{m(m-1)}(q^m-\varepsilon)\prod_{i=1}^{m-1}(q^{2i}-1)\right)\Big/(4,q^m-\varepsilon),\\
|\,\mathrm{P}\Omega(2m+1,q)| &= \left(q^{m^2}\prod_{i=1}^{m}(q^{2i}-1)\right)\Big/(2,q-1).
\end{aligned}$$

Proof For q odd, have already shown that the order of $\mathrm{SO}(n,q)$ is given by the expression in parentheses. We divide by 2 on passing to $\Omega(n,q)$, and another 2 on factoring out the scalars if and only if 4 divides $q^m-\varepsilon$. For q even, $|\,\mathrm{SO}(n,q)|$ is twice the bracketed expression, and we lose the factor 2 on passing to $\Omega(n,q) = \mathrm{P}\Omega(n,q)$.

Now we note that $|\,\mathrm{P}\Omega(2m+1,q)| = |\,\mathrm{PSp}(2m,q)|$ for all m. In the case $m=1$, these groups are isomorphic, since they are both isomorphic to $\mathrm{PSL}(2,q)$. We have also seen that they are isomorphic if q is even. We will see later that they are also isomorphic if $m=2$. However, they are non-isomorphic for $m \geq 3$ and q odd. This follows from the result of the following exercise.

Exercise 6.10 Let q be odd and $m \geq 2$.

(a) The group $\mathrm{PSp}(2m,q)$ has $\lfloor m/2 \rfloor + 1$ conjugacy classes of elements of order 2.

(b) The group $\mathrm{P}\Omega(2m+1,q)$ has m conjugacy classes of elements of order 2.

Hint: if $t \in \mathrm{Sp}(2m,q)$ or $t \in \Omega(2m+1,q) = \mathrm{P}\Omega(2m+1,q)$ satisfies $t^2 = 1$, then $V = V^+ \oplus V^-$, where $vt = \lambda v$ for $v \in V^\lambda$; and the subspaces V^+ and V^- are orthogonal. Show that there are m possibilities for the subspaces V^+ and V^- up to isometry; in the symplectic case, replacing t by $-t$ interchanges these two spaces but gives the same element of $\mathrm{PSp}(2m,q)$. In the case $\mathrm{PSp}(2m,q)$, there is an additional conjugacy class arising from elements $t \in \mathrm{Sp}(2m,q)$ with $t^2 = -1$.

It follows from the Classification of Finite Simple Groups that there are at most two non-isomorphic simple groups of any given order, and the only instances where there are two non-isomorphic groups are

$$\mathrm{PSp}(2m,q) \not\cong \mathrm{P}\Omega(2m+1,q) \text{ for } m \geq 3,\ q \text{ odd}$$

and

$$PSL(3,4) \not\cong \mathrm{PSL}(4,2) \cong A_8.$$

The lecture course will not contain detailed proofs of the simplicity of $\mathrm{P}\Omega(n,q)$, but at least it is possible to see why $\mathrm{PSO}^+(2m,q)$ contains a subgroup of index 2 for q even. Recall from Chapter 3 that, for the quadratic form

$$x_1x_2 + \cdots + x_{2m-1}x_{2m}$$

of Witt index m in $2m$ variables, the flat m-spaces fall into two families $\mathcal{F}^+$ and $\mathcal{F}^-$, with the property that the intersection of two flat m-spaces has even codimension in each if they belong to the same family, and odd codimension otherwise. Any element of the orthogonal group must fix or interchange the two families. Now, for q even, $\mathrm{SO}^+(2m,q)$ contains an element which interchanges the two families: for example, the transformation which interchanges the coordinates x_1 and x_2 and fixes all the others. So $\mathrm{SO}^+(2m,q)$ has a subgroup of index 2 fixing the two families, which is $\Omega^+(2m,q)$. (In the case where q is odd, such a transformation has determinant -1.)

7 Klein correspondence and triality

The orthogonal groups in dimension up to 8 have some remarkable properties. These include, in the finite case,

(a) isomorphisms between classical groups:

- $\mathrm{P}\Omega^-(4,q) \cong \mathrm{PSL}(2,q^2)$,
- $\mathrm{P}\Omega(5,q) \cong \mathrm{PSp}(4,q)$,
- $\mathrm{P}\Omega^+(6,q) \cong \mathrm{PSL}(4,q)$,
- $\mathrm{P}\Omega^-(6,q) \cong \mathrm{PSU}(4,q)$;

(b) unexpected outer automorphisms of classical groups:

- an automorphism of order 2 of $\mathrm{PSp}(4,q)$ for q even,
- an automorphism of order 3 of $\mathrm{P}\Omega^+(8,q)$;

(c) further simple groups:

- Suzuki groups;
- the groups $G_2(q)$ and ${}^3D_4(q)$;
- Ree groups.

In this section, we look at the geometric algebra underlying some of these phenomena.

Notation: we use $\mathrm{O}^+(2mF)$ for the isometry group of the quadratic form of Witt index m on a vector space of rank $2m$ (extending the notation over finite fields introduced earlier). We call this quadratic form Q *hyperbolic*. Moreover, the flat subspaces of rank 1 for Q are certain points in the corresponding projective space $\mathrm{PG}(2m-1,F)$; the set of such points is called a *hyperbolic quadric* in $\mathrm{PG}(2m-1,F)$.

We also denote the orthogonal group of the quadratic form

$$Q(x_1,\ldots,x_{2m+1}) = x_1x_2 + \cdots + x_{2m-1}x_{2m} + x_{2m+1}^2$$

by $\mathrm{O}(2m+1,F)$, again in agreement with the finite case.

7.1 Klein correspondence

The *Klein correspondence* relates the geometry of the vector space $V = F^4$ of rank 4 over a field F with that of a vector space of rank 6 over F carrying a quadratic form with Witt index 3.

It works as follows. Let W be the space of all 4×4 skew-symmetric matrices over F. Then W has rank 6: the above-diagonal elements of such a matrix may be chosen freely, and then the matrix is determined.

On the vector space W, there is a quadratic form Q given by

$$Q(X) = \mathrm{Pf}(X) \quad \text{for all } X \in W.$$

Recall the Pfaffian from Section 4.1, where we observed in particular that, if $X = (x_{ij})$, then

$$\mathrm{Pf}(X) = x_{12}x_{34} - x_{13}x_{24} + x_{14}x_{23}.$$

In particular, W is the sum of three hyperbolic planes, and the Witt index of Q is 3. There is an action ρ of $\mathrm{GL}(4,F)$ on W given by the rule

$$\rho(P) : X \mapsto P^\top X P$$

for $P \in \mathrm{GL}(4,F)$, $X \in W$. Now

$$\mathrm{Pf}(PXP^\top) = \det(P)\,\mathrm{Pf}(X),$$

so $\rho(P)$ preserves Q if and only if $\det(P) = 1$. Thus $\rho(\mathrm{SL}(4,F)) \le \mathrm{O}(Q)$, and since $\mathrm{SL}(4,F)$ is equal to its derived group we have $\rho(\mathrm{SL}(4,F)) \le \Omega^+(6,F)$. It is easily checked that the kernel of ρ consists of scalars; so in fact we have $\mathrm{PSL}(4,F) \le \mathrm{P}\Omega^+(6,F)$.

A calculation shows that in fact equality holds here. (More on this later.)

Theorem 7.1 $\mathrm{P}\Omega^+(6,F) \cong \mathrm{PSL}(4,F)$. ■

Examining the geometry more closely throws more light on the situation. Since the Pfaffian is the square root of the determinant, we have

$$Q(X) = 0 \text{ if and only if } X \text{ is singular.}$$

Now a skew-symmetric matrix has even rank; so if $Q(X) = 0$ but $X \neq 0$, then X has rank 2.

Exercise 7.1 Any skew-symmetric $n \times n$ matrix of rank 2 has the form

$$X(v,w) = v^\top w - w^\top v$$

for some $v, w \in F^n$.

Hint: Let B be such a matrix and let v and w span the row space of B. Then $B = x^\top v + y^\top w$ for some vectors x and y. Now by transposition we see that $\langle x, y \rangle = \langle v, w \rangle$. Express x and y in terms of v and w, and use the skew-symmetry to determine the coefficients up to a scalar factor.

Now $X(v,w) \neq 0$ if and only if v and w are linearly independent. If this holds, then the row space is spanned by v and w. Moreover,

$$X(av+cw, bc+dw) = (ad-bc)X(v,w).$$

So there is a bijection between the rank 2 subspaces of F^4 and the flat subspaces of W of rank 1. In terms of projective geometry, we have:

Proposition 7.2 *There is a bijection between the lines of* $\mathrm{PG}(3,F)$ *and the points on the hyperbolic quadric in* $\mathrm{PG}(5,F)$*, which intertwines the natural actions of* $\mathrm{PSL}(4,F)$ *and* $\mathrm{P\Omega}^+(6,F)$. ■

This correspondence is called the *Klein correspondence*, and the quadric is often referred to as the *Klein quadric*.

Now let A be a non-singular skew-symmetric matrix. The stabiliser of A in $\rho(\mathrm{SL}(4,F))$ consists of all matrices P such that $PAP^\top = A$. These matrices comprise the symplectic group (see the exercise below). On the other hand, A is a vector of W with $Q(A) \neq 0$, and so the stabiliser of A in the orthogonal group is the 5-dimensional orthogonal group on $A^\perp$ (where orthogonality is with respect to the bilinear form obtained by polarising Q). Thus, we have

Theorem 7.3 $\mathrm{P\Omega}(5,F) \cong \mathrm{PSp}(4,F)$. ■

Exercise 7.2 Let A be a non-singular skew-symmetric 4×4 matrix over a field F. Prove that the following assertions are equivalent, for any vectors $v, w \in F^4$:

(a) $X(v,w) = v^\top w - w^\top v$ is orthogonal to A, with respect to the bilinear form obtained by polarising the quadratic form $Q(X) = \mathrm{Pf}(X)$;

(b) v and w are orthogonal with respect to the symplectic form with matrix $A^\dagger$, that is, $vA^\dagger w^\top = 0$.

Here the matrices A and $A^\dagger$ are given by

$$A = \begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{pmatrix}, \qquad A^\dagger = \begin{pmatrix} 0 & a_{34} & -a_{24} & a_{23} \\ -a_{34} & 0 & a_{14} & -a_{13} \\ a_{24} & -a_{14} & 0 & a_{12} \\ -a_{23} & a_{13} & -a_{12} & 0 \end{pmatrix}.$$

Now show that the transformation induced on W by a 4×4 matrix P fixes A if and only if $PA^\dagger P^\top = A^\dagger$, in other words, P is symplectic with respect to $A^\dagger$.

Note that, if A is the matrix of the standard symplectic form, then so is $A^\dagger$.

Now, we have two isomorphisms connecting the groups $\mathrm{PSp}(4,F)$ and $\mathrm{P}\Omega(5,F)$ in the case where F is a perfect field of characteristic 2. We can apply one and then the inverse of the other to obtain an automorphism of the group $\mathrm{PSp}(4,F)$. Now we show geometrically that it must be an outer automorphism.

The isomorphism in the preceding section was based on taking a vector space of rank 5 and factoring out the radical Z. Recall that, on any coset $Z+u$, the quadratic form takes each value in F precisely once; in particular, there is a unique vector in each coset on which the quadratic form vanishes. Hence there is a bijection between vectors in F^4 and vectors in F^5 on which the quadratic form vanishes. This bijection is preserved by the isomorphism. Hence, under this isomorphism, the stabiliser of a point of the symplectic polar space is mapped to the stabiliser of a point of the orthogonal polar space.

Now consider the isomorphism given by the Klein correspondence. Points on the Klein quadric correspond to lines of $\mathrm{PG}(3,F)$, and it can be shown that, given a non-singular matrix A, points of the Klein quadric orthogonal to A correspond to flat lines with respect to the corresponding symplectic form on F^4. In other words, the isomorphism takes the stabiliser of a line (in the symplectic space) to the stabiliser of a point (in the orthogonal space).

So the composition of one isomorphism with the inverse of the other interchanges the stabilisers of points and lines of the symplectic space, and so is an outer automorphism of $\mathrm{PSp}(4,F)$.

7.2 The Suzuki groups

In certain cases, we can choose the outer automorphism such that its square is the identity. Here is a brief account.

Theorem 7.4 *Let F be a perfect field of characteristic 2. Then the polar space defined by a symplectic form on F^4 itself has a polarity if and only if F has an automorphism σ satisfying $\sigma^2 = 2$, where 2 denotes the automorphism $x \mapsto x^2$ of F.*

Proof We take the standard symplectic form

$$B((x_1,x_2,x_3,x_4),(y_1,y_2,y_3,y_4)) = x_1y_2 + x_2y_1 + x_3y_4 + x_4y_3.$$

The Klein correspondence takes the line spanned by the two points (x_1,x_2,x_3,x_4) and (y_1,y_2,y_3,y_4) to the point wit coordinates z_{ij}, for $1 \le i < j \le 4$, where $z_{ij} = x_iy_j + x_jy_i$. This point lies on the Klein quadric with equation

$$z_{12}z_{34} + z_{13}z_{24} + z_{14}z_{23} = 0,$$

and also (if the line is flat) on the hyperplane $z_{12} + z_{34} = 0$. This hyperplane is orthogonal to the point p with $z_{12} = z_{34} = 1$, $z_{ij} = 0$ otherwise. Using coordinates $(z_{13}, z_{24}, z_{14}, z_{23})$ in $p^\perp/p$, we obtain a point of the symplectic space representing the line. This gives the duality δ previously defined.

Now take a point $q = (a_1,a_2,a_3,a_4)$ of the original space, and calculate its image under the duality, by choosing two flat lines through q, calculating their images, and taking the line joining them. Assuming that a_1 and a_4 are non-zero, we can use the lines joining q to the points $(a_1,a_2,0,0)$ and $(0,a_4,a_1,0)$; their 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 compute the image of the line joining these two points, which turns out to be $(a_1^2, a_2^2, a_3^2, a_4^2)$. In all other cases, the result is the same. So $\delta^2 = 2$.

If there is a field automorphism σ such that $\sigma^2 = 2$, then $\delta\sigma^{-1}$ is a duality whose square is the identity, that is, a polarity.

Conversely, suppose that there is a polarity τ. Then $\delta\tau$ is a collineation, hence a product of a linear transformation and a field automorphism, say $\delta\tau = g\sigma$. Since $\delta^2 = 2$ and $\tau^2 = 1$, we have that $\sigma^2 = 2$ as required. ∎

It can further be shown that the set of collineations which commute with this polarity is a group G which acts doubly transitively on the set Ω of absolute points of the polarity, and that Ω is an *ovoid* (that is, each flat line contains a unique point of Ω. If $|F| > 2$, then the derived group of G is a simple group, the *Suzuki group* $\mathrm{Sz}(F)$.

The finite field $\mathrm{GF}(q)$, where $q = 2^m$, has an automorphism σ satisfying $\sigma^2 = 2$ if and only if m is odd (in which case, $2^{(m+1)/2}$ is the required automorphism). In this case we have $|\Omega| = q^2 + 1$, and $|\mathrm{Sz}(q)| = (q^2+1)q^2(q-1)$. For $q = 2$, the Suzuki group is not simple, being isomorphic to the Frobenius group of order 20.

7.3 Clifford algebras and spinors

We saw earlier (Proposition 3.11) that, if Q is a hyperbolic quadratic form on F^{2m}, then the maximal flat subspaces for Q fall into two families $\mathcal{S}^+$ and $\mathcal{S}^-$, such that if S and T are maximal flat subspaces, then $S \cap T$ has even codimension in S and T if and only if S and T belong to the same family.

In this section we represent the maximal flat subspaces as points in a larger projective space, based on the space of *spinors*. The construction is algebraic. First we briefly review facts about multilinear algebra.

Let V be a vector space over a field F, with rank m. The *tensor algebra* of V, written $\bigotimes V$, is the largest associative algebra generated by V in a linear fashion. In other words,

$$\bigotimes V = \bigoplus_{n \geq 0} \bigotimes^n V,$$

where, for example, $\bigotimes^2 V = V \otimes V$ is spanned by symbols $v \otimes w$, with $v, w \in V$, subject to the relations

$$\begin{array}{rcl} (v_1 + v_2) \otimes w & = & v_1 \otimes w + v_2 \otimes w, \\ v \otimes (w_1 + w_2) & = & v \otimes w_1 + v \otimes w_2, \\ (cv) \otimes w = \ c(v \otimes w) & = & v \otimes cw. \end{array}$$

(Formally, it is the quotient of the free associative algebra over F with basis V by the ideal generated by the differences of the left and right sides of the above identities.) The algebra is $\mathbb{N}$-graded, that is, it is a direct sum of components $V_n = \bigotimes^n V$ indexed by the natural numbers, and $V_{n_1} \otimes V_{n_2} \subseteq V_{n_1+n_2}$.

If $(e_1, \ldots, e_m)$ is a basis for V, then a basis for $\bigotimes^n V$ consists of all symbols

$$e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_n},$$

for $i_1, \ldots, i_n \in \{1, \ldots, m\}$; thus,

$$\mathrm{rk}(\bigotimes^n V) = m^n.$$

The *exterior algebra* of V is similarly defined, but we add an additional condition, namely $v \wedge v = 0$ for all $v \in V$. (In this algebra we write the multiplication as $\wedge$.) Thus, the exterior algebra $\bigwedge V$ is the quotient of $\bigotimes V$ by the ideal generated by $v \otimes v$ for all $v \in V$.

In the exterior algebra, we have $v \wedge w = -w \wedge v$. For

$$0 = (v+w) \wedge (v+w) = v \wedge v + v \wedge w + w \wedge v + w \wedge w,$$

and the first and fourth terms on the right are zero. This means that, in any expression $e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_n}$, we can rearrange the factors (possibly changing the signs), and if two adjacent factors are equal then the product is zero. Thus, the nth component $\bigwedge^n V$ has a basis consisting of all symbols

$$e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_n}$$

where $i_1 < i_2 < \ldots < i_n$. In particular,

$$\mathrm{rk}(\bigwedge^n V = \binom{m}{n},$$

so that $\bigwedge^n V + \{0\}$ for $n > m$; and

$$\mathrm{rk}(\bigwedge V) = \sum_{n=0}^{m} \binom{m}{n} = 2^m.$$

Note that $\mathrm{rk}(\bigwedge^m V) = 1$. Any linear transformation T of V induces in a natural way a linear transformation on $\bigotimes^n V$ or $\bigwedge^n V$ for any n. In particular, the transformation $\bigwedge^m T$ induced on $\bigwedge^m V$ is a scalar, and this provides a coordinate-free definition of the determinant:

$$\det(T) = \bigwedge^m T.$$

Now let Q be a quadratic form on V. We define the *Clifford algebra* $C(Q)$ of Q to be the largest associative algebra generated by V in which the relation

$$v \cdot v = Q(v)$$

holds. (We use $\cdot$ for the multiplication in $C(Q)$. Note that, if Q is the zero form, then $C(Q)$ is just the exterior algebra. If B is the bilinear form obtained by polarising Q, then we have

$$v \cdot w + w \cdot v = B(v, w).$$

This follows because

$$Q(v+w) = (v+w) \cdot (v+w) = v \cdot v + v \cdot w + w \cdot v + w \cdot w$$

and also

$$Q(v+w) = Q(v) + Q(w) + B(v,w).$$

Now, when we arrange the factors in an expression like

$$e_{i_1} \cdot e_{i_2} \cdots e_{i_n},$$

we obtain terms of degree $n-2$ (and hence $n-4, n-6, \ldots$ as we continue). So again we can say that the nth component has a basis consisting of all expressions

$$e_{i_1} \cdot e_{i_2} \cdots e_{i_n},$$

where $i_1 < i_2 < \ldots < i_n$, so that $\mathrm{rk}(C(Q)) = 2^m$. But this time the algebra is not graded but only $\mathbb{Z}_2$-graded. That is, if we let C^0 and C_1 be the sums of the components of even (resp. odd) degree, then $C^i \cdot C^j \subseteq C^{i+j}$, where the superscripts are taken modulo 2.

Suppose that Q polarises to a non-degenerate bilinear form B. Let $G = \mathrm{O}(Q)$ and $C = C(Q)$. The *Clifford group* $\Gamma(Q)$ is defined to be the group of all those units $s \in C$ such that $s^{-1}Vs = V$. Note that $\Gamma(Q)$ has an action χ on V by the rule

$$s : v \mapsto s^{-1}vs.$$

Proposition 7.5 *The action χ of $\Gamma(Q)$ on V is orthogonal.*

Proof

$$Q(s^{-1}vs) = (s^{-1}vs)^2 = s^{-1}v^2s = s^{-1}Q(v)s = Q(v),$$

since $Q(v)$, being a scalar, lies in the centre of C. ■

We state without proof:

Proposition 7.6 *(a)* $\chi(\Gamma(Q)) = \mathrm{O}(Q)$;

(b) $\ker(\chi)$ *is the multiplicative group of invertible central elements of* $C(Q)$. ■

The structure of $C(Q)$ can be calculated in special cases. The one which is of interest to us is the following:

Theorem 7.7 *Let Q be hyperbolic on F^{2m}. Then $C(Q) \cong \mathrm{End}(S)$, where S is a vector space of rank 2^m over F called the* space of spinors. *In particular, $\Gamma(Q)$ has a representation on S, called the* spin representation.

The theorem follows immediately by induction from the following lemma:

Lemma 7.8 *Suppose that $Q = Q' + yz$, where y and z are variables not occurring in Q'. Then $C(Q) \cong M_{2\times 2}(C(Q'))$.*

Proof Let $V = V' \perp \langle e, f\rangle$. Then V' generates $C(Q')$, and V', e, f generate $C(Q)$. We represent these generators by 2×2 matrices over $C(Q')$ as follows:

$$\begin{aligned} v &\mapsto \begin{pmatrix} v & 0 \\ 0 & -v \end{pmatrix}, \\ e &\mapsto \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \\ f &\mapsto \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \end{aligned}$$

Some checking is needed to establish the relations. ■

Let S be the vector space affording the spin representation. If U is a flat m-subspace of V, let f_U be the product of the elements in a basis of U. (Note that f_U is uniquely determined up to multiplication by non-zero scalars; indeed, the subalgebra of $C(Q)$ generated by U is isomorphic to the exterior algebra of U.) Now it can be shown that Cf_U and f_UC are minimal left and right ideals of C. Since $C \cong \mathrm{End}(S)$, each minimal left ideal has the form $\{T : VT \subseteq X\}$ and each minimal right ideal has the form $\{T : \ker(T) \supseteq Y\}$, where X and Y are subspaces of V of dimension and codimension 1 respectively. In particular, a minimal left ideal and a minimal right ideal intersect in a subspace of rank 1.

Thus we have a map σ from the set of flat m-subspaces of V into the set of 1-subspaces of S.

Vectors which span subspaces in the image of σ are called *pure spinors*.

Theorem 7.9 *$S = S^+ \oplus S^-$, where $\mathrm{rk}(S^+) = \mathrm{rk}(S^-) = 2^{m-1}$. Moreover, any pure spinor lies in either S^+ or S^- according as the corresponding maximal flat subspace lies in $\mathcal{S}^+$ or $\mathcal{S}^-$.* ■

Furthermore, it is possible to define a quadratic form γ on S, whose corresponding bilinear form β is non-degenerate, so that the following holds:

- if m is odd, then S^+ and S^- are flat subspaces for γ, and β induces a non-degenerate pairing between them;

- if m is even, then S^+ and S^- are orthogonal with respect to β, and γ is non-degenerate hyperbolic on each of them.

We now look briefly at the case $m = 3$. In this case, $\mathrm{rk}\, S^+ = \mathrm{rk}(S^-) = 4$. The Clifford group has a subgroup of index 2 fixing S^+ and S^-, and inducing dual representations of $\mathrm{SL}(4, F)$ on them. We have here the Klein correspondence in another form.

This case $m = 4$ is even more interesting, as we see in the next section.

7.4 Triality

Suppose that, in the notation of the preceding section, $m = 4$. That is, Q is a hyperbolic quadratic form on $V = F^8$, and the spinor space S is the direct sum of two subspaces S^+ and S^- of rank 8, each carrying a hyperbolic quadratic form of rank 8. So each of these two spaces is isomorphic to the original space V. There is an isomorphism τ (the *triality map*) of order 3 which takes V to S^+ to S^- to V, and takes Q to $\gamma|S^+$ to $\gamma|S^-$ to Q. Moreover, τ induces an outer automorphism of order 3 of the group $\mathrm{P}\Omega^+(8, F)$.

Moreover, we have:

Proposition 7.10 *A vector $s \in S$ is a pure spinor if and only if*

(a) $s \in S^+$ *or* $s \in S^-$*; and*

(b) $\gamma(s) = 0$. ■

Hence τ takes the stabiliser of a point to the stabiliser of a maximal flat subspace in $\mathcal{S}^+$ to the stabiliser of a maximal flat subspace in $\mathcal{S}^-$ back to the stabiliser of a point.

It can be shown that the centraliser of τ in the orthogonal group is the group $G_2(F)$, an *exceptional group of Lie type*, which is the automorphism group of an octonion algebra over F.

Further references for this chapter are in C. Chevalley, *The Algebraic Theory of Spinors and Clifford Algebras* (Collected Works Vol. 2), Springer, 1997.

8 Further topics

The main topic in this section is *Aschbacher's Theorem*, which describes the subgroups of the classical groups. First, there are two preliminaries: the *O'Nan–Scott Theorem*, which does a similar job for the symmetric and alternating groups; and the structure of *extraspecial p-groups*, which is an application of some of the earlier material and also comes up unexpectedly in Aschbacher's Theorem.

8.1 Extraspecial p-groups

An *extraspecial p-group* is a p-group (for some prime p) having the property that its centre, derived group, and Frattini subgroup all coincide and have order p. Otherwise said, it is a non-abelian p-group P with a normal subgroup Z such that $|Z| = p$ and P/Z is elementary abelian.

For example, of the five groups of order 8, two (the dihedral and quaternion groups) are extraspecial; the other three are abelian.

Exercise 8.1 Prove that the above conditions are equivalent.

Theorem 8.1 *An extraspecial p-group has order p^m, where m is odd and greater than* 1*. For any prime p and any odd $m > 1$, there are up to isomorphism exactly two extraspecial p-groups of order p^m.*

Proof We translate the classification of extraspecial p-groups into geometric algebra. First, note that such a group is nilpotent of class 2, and hence satisfies the following identities:

$$\begin{aligned} [xy,z] &= [x,z][y,z], && (2)\\ (xy)^n &= x^n y^n [y,x]^{n(n-1)/2}. && (3) \end{aligned}$$

(Here $[x,y] = x^{-1}y^{-1}xy$.)

Exercise 8.2 Prove that these equations hold in any group which is nilpotent of class 2.

Let P be extraspecial with centre Z. Then Z is isomorphic to the additive group of $F = \mathrm{GF}(p)$; we identify Z with F. Also, P/Z, being elementary abelian, is isomorphic to the additive group of a vector space V over F; we identify P/Z with V.

Of course, we have to be prepared to switch between additive and multiplicative notation.

The structure of P is determined by two functions $B : V \times V \to F$ and $Q : V \to F$, defined as follows. Since P/Z is elementary abelian, the commutator of any two elements of P, or the pth power of any element of P, lie in Z. So commutation and pth power are maps from $P \times P$ to F and from P to F. Each is unaffected by changing its argument by an element of Z, since

$$[xz, y] = [x, y] = [x, yz] \text{ and } (xz)^p = x^p$$

for $z \in Z$. So we have induced maps $P/Z \times P/Z \to Z$ and $P/Z \to Z$, which (under the previous identifications) are our required B and Q.

Exercise 8.3 Show that the structure of P can be reconstructed uniquely from the field F, the vector space V, and the maps B and Q above.

Now Equation (2) shows that B is bilinear. Since $[x, x] = 1$ for all x, it is alternating. Elements of its radical lie in the centre of P, which is Z by assumption; so B is nondegenerate. Thus B is a symplectic form.

In particular, $n = \mathrm{rk}(V)$ is even; so $|P| = p^m$ where $m = 1 + n$ is odd, proving the first part of the theorem.

Now the analysis splits into two cases, according as $p = 2$ or p is odd.

Case $p = 2$ Now consider the map Q. Since $|Z| = 2$, we have $[y, x] = [x, y]^{-1} = [x, y]$ for all x, y. Now Equation (3) for $n = 2$, in additive notation, asserts that

$$Q(x + y) = Q(x) + Q(y) + B(x, y),$$

In other words, Q is a quadratic form which polarises to B.

Since there are just two inequivalent quadratic forms, there are just two possible groups of each order up to isomorphism.

Case p **odd** The difference is caused by the behaviour of $p(p-1)/2$ mod p: for p odd, p divides $p(p-1)/2$. Hence Equation (3) asserts

$$Q(x + y) = Q(x) + Q(y).$$

In other words, Q is linear. Any linear function can be uniquely represented as $Q(x) = B(x, a)$ for some vector $a \in V$. Since the symplectic group has just two orbits on V, namely $\{0\}$ and the set of all non-zero vectors, there are again just two different groups. Note that the choice $a = 0$ gives a group of exponent p, while $a \neq 0$ gives a group of exponent p^2. ∎

Corollary 8.2 *(a) The outer automorphism groups of the extraspecial 2-groups of order 2^{1+2r} are the orthogonal groups $\Omega^{\varepsilon}(2r,2)$, for $3epsilon = \pm 1$.*

(b) Let p be odd. The outer automorphism group of the extraspecial p-group of order p^{1+2r} and exponent p is the general symplectic group $\mathrm{GSp}(2r,p)$ *consisting of linear maps preserving the symplectic form up to a scalar factor. The automorphism group of the extraspecial p-group of order p^{1+2r} and exponent p^2 is the stabiliser of a non-zero vector in the general symplectic group.*

Exercise 8.4 (a) Let P_1 and P_2 be groups and θ an isomorphism between central subgroups Z_1 and Z_2 of P_1 and P_2. The *central product* $P_1 \circ P_2$ of P_1 and P_2 with respect to θ is the factor group

$$(P_1 \times P_2)/\{(z^{-1}, z\theta) : z \in Z_1\}.$$

Prove that the central product of extraspecial p-groups is extraspecial, and corresponds to taking the orthogonal direct sum of the corresponding vector spaces with forms.

(b) Hence prove that any extraspecial p-group of order p^{1+2r} is a central product of r extraspecial groups of order p^3 where

- if $p = 2$, all or all but one of the factors is dihedral;
- if p is odd, all or all but one of the factors has exponent p.

We conclude with one more piece of information about extraspecial groups. Let p be extraspecial of order p^{1+2r}. The p elements of the centre lie in conjugacy classes of size 1; all other conjugacy classes have size p, so there are $p^{2r} + p - 1$ conjugacy classes. Hence there are the same number of irreducible characters. But P/P' has order p^{2r}, so there are p^{2r} characters of degree 1. It is easy to see that the remaining $p-1$ characters each have degree p^r; they are distinguished by the values they take on the centre of P.

For $p = 2$, there is only one non-linear character, which is fixed by outer automorphisms of P. Thus the representation of P lifts to the extension $P.\Omega^{\varepsilon}(2r,2)$.

For $p = 2$, suppose that P has exponent p. The subgroup $\mathrm{Sp}(2r, p)$ of the outer automorphism group acts trivially on the centre, so fixes the $p-1$ non-linear representations; again, these representations lift to $P.\mathrm{Sp}(2r, p)$.

In the case of the last remark, the representation of $P.\mathrm{Sp}(2r, p)$ can be written over $\mathrm{GF}(l)$ (l a prime power) provided that this field contains primitive pth roots of unity, that is, $l \equiv 1 \pmod p$. For the corresponding case with $p = 2$, we require primitive 4th roots of unity, that is, $l \equiv 1 \pmod 4$.

Thus, if these conditions hold, then $\mathrm{GL}(p^r, l)$ contains a subgroup isomorphic to $P.\mathrm{Sp}(2r, p)$ or $P.\Omega^{\varepsilon}(2r, 2)$ (for $p = 2$).

8.2 The O'Nan–Scott Theorem

The O'Nan–Scott Theorem for subgroups of symmetric and alternating groups is a slightly simpler prototype for Aschbacher's Theorem

A group G is called *almost simple* if $S \le G \le \mathrm{Aut}(S)$ for some non-abelian finite simple group S.

We define five classes of subgroups of the symmetric group S_n as follows:

$\mathcal{C}_1$	$\{S_k \times S_l : k + l = n, k, l > 1\}$	intransitive
$\mathcal{C}_2$	$\{S_k \wr S_l : kl = n, k, l \ge 2\}$	imprimitive
$\mathcal{C}_3$	$\{S_k \wr S_l : k^l = n, k, l \ge 2\}$	product action
$\mathcal{C}_4$	$\{\mathrm{AGL}(d, p) : p^d = n\}$	affine
$\mathcal{C}_5$	$\{(T^k).(\mathrm{Out}(T) \times S_k) : k \ge 2\}$	diagonal

In the last row of the table, T is a non-abelian simple group, and the group in question has its *diagonal action*: the stabiliser of a point is $\mathrm{Aut}(T) \times S_k = (T_d).(\mathrm{Out}(T) \times S_k)$, where the embedding of T_d in T^k is the diagonal one, as

$$T_d = \{(t, t, \ldots, t) : t \in T\},$$

and the action of $T = T_d$ is by inner automorphisms.

Now we can state the O'Nan–Scott Theorem.

Theorem 8.3 *Let G be a subgroup of S_n or A_n, not equal to S_n or A_n. Then either*

(a) G is contained in a subgroup belonging to one of the classes $\mathcal{C}_i$, $i = 1, \ldots, 5$; or

(b) G is primitive and almost simple.

Note that the action of G in case (b) is not specified.

We sketch a proof of the theorem. If G is intransitive, then it is contained in a maximal intransitive subgroup, which belongs to $\mathcal{C}_1$. If G is transitive but imprimitive, then it is contained in a maximal imprimitive subgroup, which belongs to $\mathcal{C}_2$. So we may suppose that G is primitive.

Let N be the *socle* of G, the product of its minimal normal subgroups. It is well known and easy to prove that a primitive group has at most two minimal normal subgroups; if there are two, then they are abelian. So N is a product of isomorphic simple groups.

Now the steps required to complete the proof are as follows:

- If N is abelian, then it is elementary abelian of order p^d for some prime p, and N is regular, so $n = p^d$. Then $G \leq \mathrm{AGL}(d,p) = p^d : \mathrm{GL}(d,p)$, so G is contained in a group in $\mathcal{C}_4$.

- If N is non-abelian but not simple, then it can be shown that G is contained in a group in $\mathcal{C}_3 \cup \mathcal{C}_5$.

- Of course, if N is simple, then G is almost simple.

In order to understand the maximal subgroups of S_n and A_n, there are two things to do now. The theorem shows that the maximal subgroups are either in the classes $\mathcal{C}_1$–$\mathcal{C}_5$ or almost simple. First, we must resolve the question of which of these groups contains another; this has been done by Liebeck, Praeger and Saxl. Second, we must understand how almost simple groups act as primitive permutation groups; equivalently, we must understand their maximal subgroups (since a primitive action of a group is isomorphic to the action on the right cosets of a maximal subgroup).

According to the Classification of Finite Simple Groups, most of the finite simple groups are classical groups. So this leads us naturally to the question of proving a similar result for classical groups.

8.3 Aschbacher's Theorem

Aschbacher's Theorem is the required result. After a preliminary definition, we give the eight classes of subgroups, and then state the theorem.

A subgroup H of $\mathrm{GL}(n,F)$ is said to be *irreducible* if no subspace of F^n is invariant under H. We say that H is *absolutely irreducible* if, regarding elements of H as $n \times n$ matrices over F, the group they generate is an irreducible subgroup of $\mathrm{GL}(n,K)$ for any algebraic extension field K of F.

For example, the group

$$\mathrm{SO}(2,\mathbb{R}) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\}$$

is irreducible but not absolutely irreducible since, if we write it relative to the basis $(e_1 + \mathrm{i}e_2, e_1 - \mathrm{i}e_2)$, the group would be

$$\left\{ \begin{pmatrix} \mathrm{e}^{\mathrm{i}\theta} & 0 \\ 0 & \mathrm{e}^{-\mathrm{i}\theta} \end{pmatrix} \right\}.$$

Now we describe the Aschbacher classes. The examples of groups in these classes will refer particularly to the general linear groups, but the definitions apply to all the classical groups. We let V be the natural module for the classical group G.

$\mathcal{C}_1$ consists of reducible groups, those which stabilise a subspace W of V. In $\mathrm{GL}(V)$, the stabiliser of W consists of matrices which, in block form (the basis of W coming first), have shape

$$\begin{pmatrix} A & O \\ X & B \end{pmatrix},$$

where $A \in \mathrm{GL}(k,F)$, $B \in \mathrm{GL}(l,F)$ (with $k+l=n$), and X an arbitrary $l \times k$ matrix; its structure is $F^{kl} : (\mathrm{GL}(k,F) \times \mathrm{GL}(l,F))$.

Note that, in a classical group with a sesquilinear form B, if the subspace W is fixed, then so is $W \cap W^{\perp}$. So we may assume that either $W \cap W^{\perp} = \{0\}$ (so that W is non-degenerate) or $W \leq W^{\perp}$ (so that W is flat).

$\mathcal{C}_2$ consists of irreducible but imprimitive subgroups, those which preserve a direct sum decomposition

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_t,$$

where $\mathrm{rk}(V_i) = m$ and $n = mt$; elements of the group permute these subspaces among themselves. The stabiliser of the decomposition in $\mathrm{GL}(n,F)$ is $\mathrm{GL}(m,F) \wr S_t$.

$\mathcal{C}_3$ consists of *superfield groups*. That is, a group in this class is a classical group acting on $\mathrm{GF}(q^r)^m$, where $rm = n$, and it is embedded in $\mathrm{GL}(n,q)$ by restricting scalars on the vector space from $\mathrm{GF}(q^r)$ to $\mathrm{GF}(q)$. Elements of the Galois group of $\mathrm{GF}(q^r)$ over $\mathrm{GF}(q)$ are also linear. So in $\mathrm{GL}(n,q)$, a subgroup of this form has shape $\mathrm{GL}(m,q^r):C_r$. For maximality, we may take r to be prime.

In the case of the classical group, we must sometimes modify the form (by taking its trace from $\mathrm{GF}(q^r)$ to $\mathrm{GF}(q)$); this may change the type of the form.

$\mathcal{C}_4$ consists of groups which preserve a tensor product structure $V = F^{n_1} \otimes F^{n_2}$, with $n_1 n_2 = n$. The appropriate subgroup of $\mathrm{GL}(n,F)$ is the central product $\mathrm{GL}(n_1,F) \circ \mathrm{GL}(n_2,F)$. We can visualise this example most easily by taking V to be the vector space of all $n_1 \times n_2$ matrices, and letting the pair $(A,B) \in \mathrm{GL}(n_1,F) \times \mathrm{GL}(n_2,F)$ act by the rule

$$(A,B) : X \mapsto A^{-1}XB.$$

The kernel of the action is the appropriate subgroup which has to be factored out to form the central product.

$\mathcal{C}_5$ consists of *subfield groups*, that is, subgroups obtained by restricting the matrix entries to a subfield $\mathrm{GF}(q_0)$ of $\mathrm{GF}(q)$, where $q = q_0^r$ (and we may take r to be prime).

$\mathcal{C}_6$ consists of groups with extraspecial normal subgroups. We saw in the section on extraspecial groups that the group $P.\,\mathrm{Sp}(2r,p)$ or (if $p = 2$) $P.\Omega^{\varepsilon}(2r,2)$ can be embedded in $\mathrm{GL}(p^r,l)$ if p (or 4) divides $l-1$. These, together with the scalars in $\mathrm{GF}(l)$, form the groups in this class.

$\mathcal{C}_7$ consists of groups preserving tensor decompositions of the form

$$V = V_1 \otimes V_2 \otimes \cdots \otimes V_t,$$

with $\mathrm{rk}(V_i) = m$ and $n = m^t$. These are somewhat difficult to visualise!

$\mathcal{C}_8$ consists of classical subgroups. Thus, any classical group acting on F^n can occur here as a subgroup of $\mathrm{GL}(n,F)$ provided that it is not obviously non-maximal (e.g. we exclude $\Omega^{\varepsilon}(2r,q)$ for q even, since these groups are contained in $\mathrm{Sp}(2r,q)$. However, these groups would occur as class $\mathcal{C}_8$ subgroups of the symplectic group.

Now some notation for Aschbacher's Theorem. We let $X(q)$ denote a classical group over $\mathrm{GF}(q)$, and $V = \mathrm{GF}(q)^n$ its natural module. Also, $\Omega(q)$ is the normal subgroup of $X(q)$ such that $\Omega(q)$ modulo scalars is simple; and $A(q)$ is the normaliser of $X(q)$ in the group of all invertible semilinear transformations of $\mathrm{GF}(q)^n$. A bar over the name of a group denotes that we have factored out scalars. Note that $\overline{A}(q)$ is the automorphism group of $\overline{\Omega}(q)$ except in the cases $X(q) = \mathrm{GL}(n,q)$ (where there is an outer automorphism induced by duality), $X(q) = O^+(8,q)$ (where there is an outer automorphism induced by triality), and $X(q) = \mathrm{Sp}(4,q)$ with q even (where there is an outer automorphism induced by the exceptional duality of the polar space).

Theorem 8.4 *With the above notation, let $\Omega(q) \le G \le A(q)$, and suppose that H is a subgroup of G not containing $\Omega(q)$. Then either*

(a) H is contained in a subgroup in one of the classes $\mathcal{C}_1, \ldots, \mathcal{C}_8$; or

(b) H is absolutely irreducible and almost simple modulo scalars.

Kleidman and Liebeck, *The Subgroup Structure of the Finite Classical Groups*, London Mathematical Society Lecture Note Series **129**, Cambridge University Press, 1990, gives further details, including an investigation of which of the groups in the Aschbacher classes are actually maximal.

A short bibliography on classical groups

Standard books on classical groups are Artin [2], Dieudonné [14], Dickson [13] and, for a more modern account, Taylor [22]. Cameron [5] describes the underlying geometry.

Books on related topics include Cohn [10] on division rings, Gorenstein [15] for the classification of finite simple groups, the *ATLAS* [11] for properties of small simple groups (including all the sporadic groups), the *Handbook of Incidence Geometry* [4] for a detailed account of many topics including the geometry of the classical groups, Chevalley [9] on Clifford algebras, spinors and triality, and Kleidman and Liebeck [17] on subgroups of classical groups. (The last book is a detailed commentary on the theorem of Aschbacher [3], itself the culmination of a line of research commencing with Galois and continuing through Cooperstein [12] and Kantor [16]. Cameron [6] has some geometric speculations on Aschbacher's Theorem.)

Carter [8] discusses groups of Lie type (identifying many of these with classical groups). The natural geometries for the groups of Lie type are buildings: see Tits [23] for the classification of spherical buildings, and Scharlau [21] for a modern account.

The other papers in the bibliography discuss aspects of the generation, subgroups, or representations of the classical groups. The list is not exhaustive!

References

[1] E. Artin, The orders of the classical simple groups, *Comm. Pure Appl. Math.* **8** (1955), 455–472.

[2] E. Artin, *Geometric Algebra*, Interscience, New York, 1957.

[3] M. Aschbacher, On the maximal subgroups of the finite classical groups, *Invent. Math.* **76** (1984), 469-514.

[4] F. Buekenhout (ed.), *Handbook of Incidence Geometry*, Elsevier, Amsterdam, 1995.

[5] P. J. Cameron, *Projective and Polar Spaces*, QMW Maths Notes **13**, London, 1991.

[6] P. J. Cameron, Finite geometry after Aschbacher's Theorem: $\mathrm{PGL}(n,q)$ from a Kleinian viewpoint, pp. 43–61 in *Geometry, Combinatorics and Related Topics* (ed. J. W. P. Hirschfeld et al.), London Math. Soc. Lecture Notes **245**, Cambridge University Press, Cambridge, 1997.

[7] P. J. Cameron and W. M. Kantor, 2-transitive and antiflag transitive collineation groups of finite projective spaces, *J. Algebra* **60** (1979), 384–422.

[8] R. W. Carter, *Simple Groups of Lie Type*, Wiley, New York, 1972.

[9] C. Chevalley, *The Algebraic Theory of Spinors and Clifford Algebras* (Collected Works Vol. 2), Springer, Berlin, 1997.

[10] P. M. Cohn, *Skew Field Constructions*, London Math. Soc. Lecture Notes **27**, Cambridge University Press, Cambridge, 1977.

[11] 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.

[12] B. N. Cooperstein, Minimal degree for a permutation representation of a classical group, *Israel J. Math.* **30** (1978), 213–235.

[13] L. E. Dickson, *Linear Groups, with an Exposition of the Galois Field Theory*, Dover Publ. (reprint), New York, 1958.

[14] J. Dieudonné, *La Géometrie des Groupes Classiques*, Springer, Berlin, 1955.

[15] D. Gorenstein, *Finite Simple Groups: An Introduction to their Classification*, Plenum Press, New York, 1982.

[16] W. M. Kantor, Permutation representations of the finite classical groups of small degree or rank, *J. Algebra* **60** (1979), 158–168.

[17] 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.

[18] M. W. Liebeck, On the orders of maximal subgroups of the finite classical groups, *Proc. London Math. Soc.* (3) **50** (1985), 426–446.

[19] G. Malle, J. Saxl and T. Weigel, Generation of classical groups, *Geom. Dedicata* **49** (1994), 85–116.

[20] H. Mäurer, Eine Charakterisierung der Permutationsgruppe $\mathrm{PSL}(2,K)$ über einem quadratisch abgeschlossenen Körper K der Charakteristik $\neq 2$, *Geom. Dedicata* **36** (1990), 235–237.

[21] R. Scharlau, Buildings, pp. 477–645 in *Handbook of Incidence Geometry* (F. Buekenhout, ed.), Elsevier, Amsterdam, 1995.

[22] D. E. Taylor, *The Geometry of the Classical Groups*, Heldermann Verlag, Berlin, 1992.

[23] J. Tits, *Buildings of Spherical Type and Finite BN-Pairs*, Lecture Notes in Math. **382**, Springer–Verlag, Berlin, 1974.