Geometry: Finite, Real, Imaginary

Chris Godsil

March 16, 2020

Preface

This is a text on design theory and geometry.

To Do

Designs

- Need to define resolvability earlier. Perhaps in orthogonal array chapter.
- Treat lattices, add Leech lattice.

Geometry

- Completely rewrite chapter on affine geometries.
- Make terminology compatible with Shult.
- Construct graphs using sets of points at infinity in Affine chapter.
- Add section on Grassmann and bilinear forms graphs to projective spaces?
- Treat Bose-Burton explicitly.
- Add section on Möbius planes, prove that planes of even order are egg-like.
- Flocks. Elliptic flocks in odd characteristic are linear. (Bruen and Fisher.)
- Discuss Buekenhout-Shult theorem.
- No maximal arcs in odd characteristic.

Lines

- Complete material on 276 lines in $\mathbb{R}^{23}$.
- Have a chapter on lines, angles, bounds and frames over $\mathbb{R}$ and $\mathbb{C}$, followed by chapters on real and complex examples.
- Semifields, tensor cubes.

Overall

- Add bibliography.

Contents

Part I

Designs

Chapter 1

Block Designs

1.1 Incidence Structures

An *incidence structure* $(\mathcal{P}, \mathcal{B})$ consists of a set of points $\mathcal{P}$, a set of blocks $\mathcal{B}$ and an incidence relation on $\mathcal{P} \times \mathcal{B}$. Thus a point and a block are either incident or not; in the first case we may say that the point lies in the block, or that the block lies on the point. What we are calling blocks may sometimes have another name, for example, in geometry the blocks are usually called lines. The point and block sets are disjoint. It is quite common to find that the blocks are defined to be subsets of the point set, but this is not a requirement.

Any graph determines an incidence structure where the vertices are the points and the edges are the blocks. A point is incident with the edges that contain it. A planar graph determines three incidence structures. The first is the one just described. The second has the vertices as points and the faces as its blocks. The third has the blocks as points and the edges as blocks.

If $\mathcal{P}$ and $\mathcal{B}$ are the point and block sets of an incidence structure, then we may view $\mathcal{B}$ as the point set and $\mathcal{P}$ as the block set of a second incidence structure. This is called the *dual* of the first incidence structure. The dual of the dual is the original structure.

If $(\mathcal{P}, \mathcal{B})$ is an incidence structure, its *incidence graph* is the graph with vertex set $\mathcal{P} \cup \mathcal{B}$, where a pair of vertices u and v are adjacent if one is a point and the other is a block incident with it. The incidence graph is is bipartite, with bipartition $(\mathcal{P}, \mathcal{B})$. In fact we will view the incidence

graph as *bicolored*: in addition to the graph itself we are provided with the ordered pair $(\mathcal{P}, \mathcal{B})$ which specifies a 2-coloring of the bipartite graph. For example, the incidence graphs of an incidence structure and its dual have the same underlying bipartite graph, but they have different 2-colourings. The incidence graph provides a very useful tool for working with incidence structures.

An incidence structure $(\mathcal{P}, \mathcal{B})$ is *point regular* if each point is incident with the same number of blocks; it is *block regular* if each block is incident with the same number of points. (The terms "regular" and "uniform" are sometimes used.) The incidence structure formed by the vertices and edges of a loopless graph is uniform—each edge is incident with exactly two vertices—but it is regular if and only if the underlying graph is regular. An incidence structure is *thick* if the minimum valency of its incidence graph is at least three.

We will say that an incidence structure is *connected* if its incidence graph is connected. A connected bipartite graph has a unique 2-coloring, and so a connected incidence structure is determined up to duality by its incidence graph.

An incidence structure is a *partial linear space* if each pair of points lies in at most one block. It is a *linear space* if each pair of points lies on exactly one line. A *dual linear space* is an incidence structure whose dual is a linear space. (This suggests, correctly, that the dual of a linear space need not be a linear space.)

1.1.1 Lemma. *An incidence structure is a partial linear space if and only if its incidence graph has girth at least six.*

Proof. Suppose a and b are points and C and D are blocks. Then the vertices a, C, b, D form a 4-cycle in the incidence graph if and only if a and b are both incident with C and D. □

One corollary of this is that if an incidence structure is a partial linear space, then so is its dual. (As we noted above, the dual of a linear space might not be a linear space.)

Since the incidence graph of a partial linear space does not contain a copy of $K_{2,2}$ it cannot contain a copy of $K_{2,m}$ where $m \geq 2$. It follows that we cannot have two blocks incident with exactly the same set of points, and in this case it is natural to identify a block with the subset of $\mathcal{P}$ consisting of the points incident with it.

Suppose $(\mathcal{P}, \mathcal{B})$ is an incidence structure. If $a, b \in \mathcal{P}$, we may define the *line through a and b* to be the intersection of all blocks incident with both a and b. The incidence structure formed from the points and lines just constructed is a partial linear space; it is usually interesting only if it is thick.

A subset S of the points of a partial linear space is a *subspace* if any line that contains two points of S is a subset of S.

Besides the incidence graph, there are two further graphs associated with a partial linear space. The vertices of the *point graph* are the points of the incidence structure, and two points are adjacent if and only if they are distinct and collinear, that is, there is a line that contains them both. The *line graph* has the lines are vertices, and two lines are adjacent if and only if the are distinct and a point in common. (Thus the line graph is the point graph of the dual of the partial linear space.

The *incidence matrix* N of a finite incidence structure $(\mathcal{P}, \mathcal{B})$ is the 01-matrix with rows indexed by $\mathcal{P}$, columns by $\mathcal{B}$ and with $N_{x,\beta} = 1$ if and only if the point x is incident with the block β. Then N^T is the incidence matrix of the dual structure and the adjacency matrix of the incidence graph is

$$\begin{pmatrix} 0 & N \\ N^T & 0 \end{pmatrix}.$$

A *parallel class* in an incidence structure is a set of blocks that partitions the point set. For example, if X is a graph then its vertices and edges form an incidence structure and a parallel class is a perfect matching. An incidence structure is *resolvable* if we can partition its block set into parallel classes. So in our graph example, the incidence structure is resolvable if and only if the graph as a 1-factorization.

1.2 Designs

We say that an incidence structure $(V, \mathcal{B})$ has *strength* at least t if for $s = 1, \ldots, t$ there are constants $\lambda_1, \ldots, \lambda_t$ such that each s-subset of V is incident with exactly λ_s elements of B. We take the view that $\lambda_0 = b$. A *t-design* is a uniform incidence structure with strength at least t. The usual convention is the "design" by itself means "2-design". Note that the same design can, for example, be both a 2-design and 3-design. However the parameters will be different. A design is *simple* if no two blocks are

incident with the same set of points. Generally our designs will be simple, and you may assume that they are simple in the absence of any warning.

A 2-design with $\lambda = 1$ is a partial linear space. A *Steiner system* is a t-design with $\lambda_t = 1$.

We offer some simple examples of designs starting with the canonical first example, the *Fano plane*.

Here we have $V = \mathbb{Z}_7$ and the blocks are as follows:

$$\{0, 1, 3\}$$
$$\{1, 2, 4\}$$
$$\{2, 3, 5\}$$
$$\{3, 4, 6\}$$
$$\{4, 5, 0\}$$
$$\{5, 6, 1\}$$
$$\{6, 0, 2\}$$

This is a block design with parameters $(v, b, r, k, \lambda) = (7, 7, 3, 3, 1)$. If we take the complement in $\mathbb{Z}_7$ of each block of the Fano plane we get a design on 7 points with block size 4. This holds in general, i.e., if we take the complement of each block in a design we obtain another design.

In a general block design, b is very large and it may be inconvenient or impossible to present the design by listing its blocks. The Fano plane is an example of a *difference set* construction. A difference set S in an abelian group G is a subset of G with the property that each non-zero element of G appears the same number of times as a difference of two elements of S. Here $\alpha = \{0, 1, 3\}$ is a difference set for $G = \mathbb{Z}_7$. If G is an abelian group and $S \subseteq G$ then the set

$$S + g = \{x + g \mid x \in S\}$$

is called a *translate* of S. In our example, the design consists of all translates of α. The Fano plane is the projective plane of order two (which we will explain later.)

Difference set constructions are attractive, and so we present another example, Let $V = \mathbb{Z}_{11}$, then $\alpha = \{0, 2, 3, 4, 8\}$ is a difference set and the set of all translates of α is a 2-design with parameters

$$(v, b, r, k, \lambda) = (11, 11, 5, 5, 2).$$

A design with $b = v$ is called a *symmetric design*. We see later on that $k = r$ in symmetric designs.

After projective planes, the next most important family of designs come from *affine planes*. Let V be a vector space and let $\mathcal{B}$ be the set of lines (a line is a coset of a 1-dimensional subspace). This is a 2-design with $\lambda = 1$. If we take V to be the 2-dimensional vector space over $\mathbb{Z}_3$ we get a 2-(9,3,1) design with $b = 12$ and $r = 4$. So

$$V = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \mid a, b \in \mathbb{Z} \right\}.$$

This is an affine plane of order three.

Finally we consider an infinite class of uninteresting examples. A design is *trivial* if $k \in \{0, 1, v-1, v\}$. These are valid designs, although not so interesting. For a *non-trivial* design, $2 \leq k \leq v-2$. The *complete design* consists of all k-subsets of a set of size v.

1.3 Relations between Parameters

As noted above, the parameters v, b, r, k, λ are not independent. Define a *flag* in an incidence structure to be an ordered pair (v, α) where $v \in V$ and $\alpha \in \mathcal{B}$. Then by counting with respect to the first or second coordinate we obtain $r|V|$ $r\,|V|$ and $k\,|\mathcal{B}|$ respectively as the number of blocks. Thus

$$vr = kb \qquad \text{or} \qquad \frac{v}{k} = \frac{b}{r}.$$

Now take ordered triples (v, w, α) where $v \neq w$ and $v, w \in \alpha$. If we count the number of such triples with respect to the first two coordinates or with respect to the third coordinate we obtain $v(v-1)\lambda$ and $bk(k-1)$ respectively. Thus

$$v(v-1)\lambda = bk(k-1)$$

or equivalently

$$\frac{v(v-1)}{k(k-1)} = \frac{b}{\lambda}.$$

Also, since $bk = vr$, we have $v(v-1)\lambda = vr(k-1)$. Therefore

$$\frac{v-1}{k-1} = \frac{r}{\lambda}$$

We see that the ratios b/λ and r/λ are determined by v and k—this is worth remembering.

Example. Suppose $\lambda = 1$ and $k = 3$. Then $b = v(v-1)/6$ and $r = (v-1)/2$. Thus

$$v \equiv 1, 3 \qquad \mod 6.$$

So the possible parameter sets with $\lambda = 1$, $k = 3$ and $v \leq 15$ are

$$(3,3,1), (7,3,1),\ (9,3,1),\ (13,3,1),\ (15,3,1)$$

and so on. Note that for a 2-(13,3,1) design we get $b = 26$ and so the number of blocks must be twice the number of points. This observation suggests trying to construct such a design by using a difference set with two initial blocks. Designs with $\lambda = 1$ and $k = 3$ are called *Steiner triple systems*.

In general, if we have a design with strength t and $s \leq t$, we have the parameter relations

$$\binom{v}{s}\lambda_s = b\binom{k}{s}$$

or, equivalently

$$\frac{\binom{v}{s}}{\binom{k}{s}} = \frac{\lambda_0}{\lambda_s}.$$

One consequence of this is that if our design has strength t and $r, s \leq t$, then the ratio λ_s/λ_t is determined by v and k—thus it does not depend on the structure of the design.

1.4 Matrices and Maps

The *incidence matrix* of an incidence structure $(\mathcal{P}, \mathcal{B})$ has rows indexed by $\mathcal{P}$, columns indexed by $\mathcal{B}$ and its ij-th entry is 0 or 1 according as the i-th point is incident with the j-th block, or not. We see that incidence matrix depends on an ordering of the points and blocks, but this will not cause any problems. If the incidence matrix of an incidence structure is B, then the adjacency matrix of its incidence graph is

$$\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}.$$

If $\mathcal{G}$ is an incidence structure with incidence matrix B, then the incidence matrix of the dual of $\mathcal{I}$ is B^T.

Let $\mathcal{G}_1 = (\mathcal{P}_1, \mathcal{B}_1)$ and $\mathcal{G}_2 = (\mathcal{P}_2, \mathcal{P}_2)$ be two incidence structures. A *homomorphism* from $\mathcal{G}_1$ to $\mathcal{G}_2$ consists of a pair of maps π andβ such that if $a \in \mathcal{P}_1$ and $B \in \mathcal{B}_2$ and a is incident with B, then $\pi(a)$ is incident with $\beta(B)$. A homomorphism is an *isomorphism* if π and β are bijections and the pair (π^{-1}, β^{-1}) is a homomorphism from $\mathcal{G}_2$ to $\mathcal{G}_1$. The condition that the inverse maps form a homomorphism is equivalent to requiring that $\pi(a)$ and $\beta(B)$ are incident in $\mathcal{G}_2$ if and only if they are incident in $\mathcal{G}_1$. Finally, an *automorphism* of an incidence structure is an isomorphism from the incidence structure to itself. The set of automorphisms of an incidence structure forms its *automorphism group*.

We can represent automorphisms by pairs of permutations. If π is a permutation of a set S and $i \in S$, we denote the image of i under π by i^π (and sometimes even by $i\pi$.) If $(\mathcal{P}, \mathcal{B})$ is an incidence structure, and π and β are permutations of $\mathcal{P}$ and $\mathcal{B}$ respectively, then (π, β) is an automorphism if a^π is incident with B^β if and only a is incident with B.

Now suppose our incidence structure has v points and b blocks and let $e_1, \ldots, e_v$ and $f_1, \ldots, f_b$ respectively denote the standard bases of $\mathbb{R}^v$ and $\mathbb{R}^b$. If π is a permutation of the points, then there is a unique linear map of $\mathbb{R}^v$ to itself that sends e_i to $e_{i\pi}$ and the matrix that represents this linear map is a permutation matrix. The pair of permutation matrices (P, Q) determines an automorphism of our incidence structure if and only if

$$PBQ^T = B.$$

(Taking the transpose of Q here is arbitrary, but useful.) If no two blocks are incident with the same set of points, then the block permutation of an automorphism is determined by the point permutation. In this case, a permutation π of V will be an automorphism if, when $x \in V$ and $\alpha \in \mathcal{B}$, we have x is incident with β if and only if x^π is incident with α^π. (Here we get α^π by applying π to each element of α.)

Chapter 2

Symmetric Designs

2.1 Incidence Matrices of Designs

For a (v, b, r, k, λ) design there are $\lambda = \lambda_2$ blocks on each pair of points. This gives

$$\frac{v(v-1)}{k(k-1)} = \frac{b}{\lambda}, \quad \frac{v-1}{k-1} = \frac{r}{\lambda}.$$

Our design has an incidence matrix N with the following properties:

(a) $N\mathbf{1} = r\mathbf{1}$

(b) $\mathbf{1}^T N = k\mathbf{1}^T$

(c) $NN^T = (r-\lambda)I + \lambda J$

where J is the matrix with all entries equal to one. These equations hold if and only if the corresponding incidence structure is a 2-design.

2.1.1 Theorem. *If $\mathcal{D}$ is a 2-design with parameters (v, b, r, k, λ) and $\mathcal{D}$ has at least 2 points and at least 2 blocks, then $b \geq v$.*

Proof. We aim to show that the rows of N are linearly independent over $\mathbb{R}$ and, since we are working over the reals, can do this by proving that NN^T is invertible. Now

$$NN^T = (r-\lambda)I + \lambda J$$

and we can write down the inverse explicitly. The key is to note that

$$(xI + J)(yI + J) = xyI + (x+y)J + vJ = xyI + (x+y+v)J,$$

from which it follows that $xI + J$ is invertible if $x \neq 0$. □

The inequality $b \geq v$ is called Fisher's inequality. Can we have $b = v$? We've already seen two examples: the 2-$(7, 3, 1)$ and 2-$(11, 5, 2)$ designs. A design with $v = b$ is called a *symmetric design*. Note that $b = v$ if and only if $r = k$.

2.1.2 Theorem. *If $\mathcal{D}$ is a symmetric design with parameters (v, k, λ), then any two distinct blocks have exactly λ points in common.*

Proof. Since $b = v$, the incidence matrix N is invertible. We have

$$NN^T = (r - \lambda)I + \lambda J$$

The rows and columns of $N^T N$ are indexed by the blocks of the design; the α, β entry of $N^T N$ is the size of $|\alpha \cap \beta|$. We want to show that $|\alpha \cap \beta|$ is constant for $\alpha \neq \beta$ and is equal to k when $\alpha = \beta$.

$$\begin{aligned} N^T N &= N^{-1}\left(NN^T\right)N \\ &= N^{-1}\left((r - \lambda)I + \lambda J\right)N \\ &= (r - \lambda)I + \lambda N^{-1}JN \end{aligned}$$

Note that since $N\mathbf{1} = k\mathbf{1}$, we have

$$\frac{1}{k}\mathbf{1} = N^{-1}\mathbf{1}.$$

It follows that

$$\lambda N^{-1}JN = \lambda N^{-1}\mathbf{1}\mathbf{1}^T N = \lambda\frac{1}{k}k\mathbf{1}\mathbf{1}^T = \lambda J$$

and hence $N^T N = (r - \lambda)I + \lambda J$. □

2.1.3 Corollary. *The dual of a symmetric design is a symmetric design.* □

In general the dual of a block design is not a block design, as you may verify easily. We also see that if $b = v$, we have shown that $N^T N = NN^T$, i.e., the incidence matrix of a symmetric design is normal.

2.2 Constructing Symmetric Designs

Suppose that we have a symmetric design with parameter set (v, k, λ). We know that

$$\frac{v(v-1)}{k(k-1)} = \frac{b}{\lambda} = \frac{v}{\lambda}$$

so that $v = 1 + \frac{k^2-k}{\lambda}$. For example if we let $\lambda = 1$, then the following table lists the possible parameter sets when $v \leq 9$.

k	v	
2	3	
3	7	
4	13	
5	21	
6	31	
7	43	
8	57	
9	73	

The fact that there is no symmetric $(43, 7, 1)$-design follows from the Bruck-Ryser-Chowla theorem, which we will meet before long. We now present a construction for a class of symmetric designs using subspaces of vector spaces. This class will include all the designs that do exist in the above table.

Let V be a vector space of dimension d over $GF(q)$, where in all interesting cases $d \geq 3$. We build an incidence structure as follows: The points are the 1-dimensional subspaces of V and the blocks are the subspaces with dimension $d-1$. Incidence is determined by inclusion (symmetrized). We call the $d-1$ dimensional subspaces *hyperplanes*. The number of elements of V is q^d, and the number of elements in a 1-dimensional subspace is q. The 1-dimensional subspaces partition $V \setminus 0$ into sets of size $q-1$. So there are $\frac{q^d-1}{q-1}$ points in our incidence structure.

Each hyperplane is the kernel of a $1 \times d$ matrix $[a_1, ..., a_d] = a$. If a and b are two non-zero vectors of length d, then $\ker(a) = \ker(b)$ if and only if b is a non-zero, scalar multiple of a. It follows that the number of hyperplanes is equal to the number of 1-dimensional subspaces, that is, $\frac{q^d-1}{q-1}$.

The non-zero vectors $[a_1, ..., a_d]$ and $[b_1, ..., b_d]$ determine distinct hyperplanes if and only if the matrix

$$\begin{pmatrix} a_1 & ... & a_d \\ b_1 & ... & b_d \end{pmatrix}$$

has rank two. The kernel of this matrix is a subspace of dimension $d-2$ and the 1-dimensional subspaces are the 1-dimensional subspaces that lie on both hyperplanes.

This allows us to conclude that the subspaces of dimension one and $d-1$ are the points and blocks of a symmetric design with parameters

$$\left(\frac{q^d - 1}{q-1}, \frac{q^{d-1} - 1}{q-1}, \frac{q^{d-2} - 1}{q-1} \right).$$

If $d = 3$ then these designs are symmetric and $\lambda = 1$, thus they are projective planes. However if $d \geq 4$ then the designs are new to us.

2.3 Two Open Problems

Question: *Is there a symmetric design with $\lambda = 1$ where $k-1$ is not a prime power?*

We will see that a symmetric design where $\lambda = 1$ is also known as a projective plane. The usual formulation of thhis question is to ask whether the order of a projective plane must be a prime power.

We know comparatively little about the symmetric designs where $\lambda \geq 2$. Thus we do not know the answer to the following.

Question: *Let ℓ be a fixed integer greater than 1. Do there exist infinitely many symmetric designs where $\lambda = \ell$?*

Note that the complement of a projective plane is a symmetric design with $\lambda > 1$, so there infinitely many symmetric designs where $\lambda > 1$. But there is no value of $\ell > 1$ for which we know the answer to the above question.

2.4 Symmetric Designs and Quadratic Forms

If N is the incidence matrix of a symmetric design with parameters (v, k, λ), then

$$NN^T = (k - \lambda)I + \lambda J.$$

We derive a somewhat simpler condition on an extended version of N. If we define

$$\widehat{N} := \begin{pmatrix} N & \mathbf{1} \\ \lambda \mathbf{1}^T & k \end{pmatrix}$$

then we can calculate and find that

$$\widehat{N} \begin{pmatrix} I & 0 \\ 0 & -\lambda \end{pmatrix} \widehat{N}^T = (k - \lambda) \begin{pmatrix} I & 0 \\ 0 & -\lambda \end{pmatrix}.$$

Since the determinant of the right side is not zero, we deduce that $\det(\widehat{N}) \neq 0$ and therefore $\widehat{N}$ is invertible.

If B is a symmetric matrix over a field $\mathbb{F}$ then $x^T Bx$ is a homogeneous quadratic polynomial in the entries of x, and such a polynomial is known as a quadratic form over $\mathbb{F}$. (We will discuss these in more details shortly.) If G is an invertible matrix over $\mathbb{F}$, the quadratic forms associated to B and $G^T BG$ are said to be equivalent. Hence one consequence of our previous calculation is that if a symmetric (v, k, λ)-design exists and $n = k - \lambda$, then the quadratic forms determined by the matrices

$$\begin{pmatrix} I & 0 \\ 0 & -\lambda \end{pmatrix}, \qquad n \begin{pmatrix} I & 0 \\ 0 & -\lambda \end{pmatrix}$$

are equivalent. Number theorists have worked out a useful characterization of when two quadratic forms over $\mathbb{Q}$ are equivalent, and these can be used to show that certain symmetric designs do not exist. In the following sections we present a version of this theory.

2.5 Bilinear Forms

We introduce some of the theory of bilinear and quadratic forms.

Let V be a vector space over a field $\mathbb{F}$. A *bilinear form* β on V is a function from $V \times V$ to $\mathbb{F}$ that is linear in both variables. If

$$\beta(u, v) = \beta(v, u)$$

for all u and v we say that β is *symmetric*. To get the canonical examples, take a symmetric matrix B and define

$$\beta(u, v) = u^T B v.$$

If β is a bilinear form on V and $v \in V$, then we define

$$v^\perp := \{x : \beta(v, x) = 0\}.$$

If $U \leq V$, then

$$U^\perp := \cap_{u \in U} u^\perp$$

It is possible that $v \in v^\perp$ and $v \neq 0$, but we do have that

$$\dim(U^\perp) = \dim(V) - \dim(U)$$

and $(U^\perp)^\perp = U$. We see that if $v \neq 0$ then $v^\perp$ is a subspace of V with codimension at most 1. We say β is *non-degenerate* if $v^\perp = V$ implies $v = 0$.

We could define a quadratic form in variables $x_1, \ldots, x_d$ to be a polynomial

$$\sum_{i \leq j} a_{i,j} x_i x_j,$$

where coefficients come from some field $\mathbb{F}$. In other words it is a homogeneous quadratic polynomial. However rather than define a form in terms of a representation, we adopt a definition in terms of its important proerties.

A *quadratic form* $\mathcal{Q}(v)$ over V is function from V to $\mathbb{F}$ such that

(a) $\mathcal{Q}(au) = a^2 \mathcal{Q}(u)$ for all a in $\mathbb{F}$ and $u \in V$.

(b) $\mathcal{Q}(u + v) - \mathcal{Q}(u) - \mathcal{Q}(v)$ is a symmetric bilinear form on V.

For example, if β is a symmetric bilinear form then

$$\beta(x+y, x+y) = \beta(x, x)+\beta(x, y)+\beta(y, x)+\beta(y, y) = \beta(x, x)+\beta(y, y)+2\beta(x, y)$$

and so $\mathcal{Q}_\beta(x) := \beta(x, x)$ is a quadratic form. If 2 is invertible in $\mathbb{F}$, then the quadratic form determines the bilinear form. In these notes we will restrict ourselves to forms over fields in which 2 is invertible. If A is a square matrix then $q(x) = x^T Ax$ is a quadratic form.

A *quadratic space* is a pair (V, q), where V is a vector space and q is a quadratic form on V. If U is a subspace of V then U along with the restriction $q \restriction U$ of q to U, is a quadratic space, a subspace. Next suppose that V is the direct sum of U_1 and U_2 and q_1 and q_2 are quadratic forms on U_1 and U_2 respectively. If we define

$$q(u_1, u_2) := q_1(u_1) + q_2(u_2)$$

then q is a quadratic form on V. We say that the quadratic space (V, q) is the *sum* of the spaces (U_1, q_1) and (U_2, q_2).

2.6 Radicals

The bilinear form we are most familiar with is the usual dot product on $\mathbb{R}^n$. This has the property that if $\beta(x, x) = 0$ then $x = 0$, and this property is not shared by bilinear forms in general. However this is not an insurmountable problem. The difficulty in extending the standard ideas about orthogonality to general forms arise because there may be vectors a such that $\beta(a, x) = 0$ for all x.

For a subset S of the vector space V, we define $S^\perp$ by

$$S^\perp := \{x \in V : \beta(x, a) = 0, \ \forall a \in S\}$$

It is easy to verify that $S^\perp$ is a subspace. If S itself is a subspace then

$$\dim(S^\perp) = \dim(V) - \dim(S)$$

and one consequence of this is that $(S^\perp)^\perp = S$. If a is a non-zero vector then

$$\dim(a^\perp) \geq \dim(V) - 1$$

and $\dim(a^\perp) = \dim(V)$ if and only if $\beta(a, x) = 0$ for all x in V. We define the *radical* of a symmetric bilinear form β to be the set of vectors a such that $\beta(a, x) = 0$ for all x. The radical of a form is a subspace (exercise). If

$U \leq V$ then the radical of U is $U \cap U^\perp$. We also see $a^\perp = V$ if and only if a lies in the radical of V. Note that

$$S^\perp = \bigcap_{a \in S} a^\perp.$$

A quadratic form is *non-degenerate* if the radical of its quadratic space is zero.

If $\beta(x, y) := x^T A y$, then a lies in the radical of β if and only if $a^T B y = 0$ for all y, and this holds if and only if $a^T A = 0$. Hence the radical of β is zero if and only if A is non-singular.

If $U \leq V$ and $U \cap U^\perp = 0$, then a dimension argument yields that $V = U + U^\perp$ and therefore V is direct sum of U and $U^\perp$. Recall that subspace V of W is a complement to a subspace U if

$$U + V = W, \quad U \cap V = 0.$$

Thus if the radical of the subspace U is zero, then $U^\perp$ is a complement to U.

2.7 Equivalence of Forms

We assume now that 2 is an invertible element of our underlying field. Hence any quadratic form $\mathcal{Q}$ we discuss can be written in the form (sorry)

$$\mathcal{Q}(x) = x^T A x,$$

where A is symmetric.

Two quadratic forms $\mathcal{Q}_1$ and $\mathcal{Q}_2$ in d variables over $\mathbb{F}$ are *equivalent* if there is an invertible matrix G such that, for all x and y in V

$$\mathcal{Q}_2(x) = \mathcal{Q}_1(Gx).$$

It is easy to check that this is, as the name suggests, an equivalence relation. Two symmetric matrices A and B are *congruent* if there is an invertible matrix G over $\mathbb{F}$ such that

$$B = G^T A G.$$

We write $A \approx B$ to denote that A and B are congruent. If two quadratic forms are equivalent then their associated bilinear forms are congruent.

(And the converse is true if the characteristic of the underlying field is not 2.)

If P is a permutation matrix, then A and P^TAP are equivalent, hence

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \approx \begin{pmatrix} b & 0 \\ 0 & a \end{pmatrix}.$$

Also if $c \neq 0$, then

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \approx \begin{pmatrix} c^2a & 0 \\ 0 & b \end{pmatrix}.$$

We have already seen that if a symmetric (v, k, λ)-design exists, then the matrices

$$\begin{pmatrix} I & 0 \\ 0 & -\lambda \end{pmatrix}, \quad n\begin{pmatrix} I & 0 \\ 0 & -\lambda \end{pmatrix}$$

give rise to equivalent forms.

2.8 Diagonal Forms

A quadratic form $q(x) = x^TAx$ is *diagonal* if A is a diagonal matrix. If (V, q) is a quadratic space and the bilinear form associated with q is β, we say that basis $v_1, \ldots, v_d$ for V is an *orthogonal basis* relative to q if $\beta(v_i, v_j) = 0$ when $i \neq j$. Thus if q is diagonal then the standard basis is an orthogonal basis for q.

2.8.1 Theorem. *If q is a nondegenerate quadratic form on V, then V contains an orthogonal basis relative to q.*

Proof. Since q is nondegenerate there is a vector v in V such that $q(v) \neq 0$. Hence

$$\langle v \rangle \cap v^\perp = 0$$

and we set $V_1 = v^\perp$. Note that the restriction of q to V_1 is nondegenerate. (Why?) By induction on dimension, we deduce that V_1 contains an orthogonal basis, and this basis extended by v is an orthogonal basis for V. □

2.8.2 Corollary. *Any non-degenerate quadratic form over a field of characteristic different from two is equivalent to a diagonal form.*

Proof. We continue with the notation of the theorem. Suppose $q(x) = x^T Ax$. If G is the linear map that maps e_r to the r-th member of this basis, for each r, then $G^T AG$ is diagonal. □

2.9 Isometries

Two quadratic spaces (V_1, q_1) and (V_2, q_2) *isometric* if there is an invertible linear map $L : V_1 \to V_2$ such that

$$q_2(Lv) = q_1(v)$$

for all v. We see that two forms q_1 and q_2 on V are equivalent if and only if the associated quadratic spaces (V, q_1) and (V, q_2) are congruent.

2.9.1 Lemma. *Let (V, q) be a quadratic space and suppose u and v are elements of V such that $q(u) = q(v) \neq 0$. Then there is an isometry of V that maps u to v.*

Proof. First we define a class of isometries. Let β be the symmetric bilinear form associated with q. If $a \in V$ and $q(a) \neq 0$, define the map τ_a on V by

$$\tau_a(v) := v - 2\frac{\beta(a, v)}{q(a)}a.$$

Then τ_a is linear and τ_a^2 is the identity. You may also check that $q(\tau_a(v)) = q(v)$ for all v; whence τ_a is an isometry.

Your next exercise is to show that if $q(u - v) \neq 0$, then τ_{u-v} swaps u and v.

Now suppose that $q(u) = q(v) \neq 0$. If $q(u - v) \neq 0$ and $a = u - v$, then $\tau(a)$ swaps u and v. If $q(u + v) \neq 0$ and $b = u + v$, then $\tau(b)$ swaps u and $-v$, and therefore $-\tau_b$ swaps u and v. If the worst happens and

$$q(u - v) = q(u + v) = 0$$

then

$$0 = q(u - v) + q(u + v) = 2q(u) + 2q(v) = 4q(u) \neq 0.$$

This proves the lemma. □

2.10 Cancellation

In terms of quadratic spaces, our goal is to prove that if U_0, U_1 and U_2 are quadratic spaces and

$$U_0 \oplus U_1 \cong U_0 \oplus U_2$$

then $U_1 \cong U_2$. In terms of quadratic forms we want to show that if q_0, q_1 and q_2 are quadratic forms in three disjoint sets of variables and $q_0 + q_1$ is equivalent to $q_0 + q_2$, then q_1 and q_2 are equivalent.

2.10.1 Theorem. *Suppose U_1 and U_2 are non-zero subspaces of the quadratic space (V, q) and the radical of U_1 is zero. Then if there is an isometry $\rho : U_1 \to U_2$, there is an isometry from V to itself whose restriction to U_1 is equal to ρ.*

Proof. If q vanishes on U_1 then the radical of U_1 is U_1, so we see that there is a vector u in U_1 such that $q(u) \neq 0$. By Lemma 2.9.1, there is an isometry σ on V such that $\sigma(\rho(u)) = u$. If $\dim(U_1) = 1$, we are done—we can take σ^{-1} to be the required isometry of V.

So $\sigma\rho$ is an isometry from U_1 to $\sigma(U_2)$ that fixes u. If $\sigma\rho$ extends to an isometry τ (say) of V, then τ followed by σ^{-1} is an isometry of V that extends ρ.

We proceed by induction on $\dim(U_1)$. Now U_1 is the sum of the span of u and the space $u^\perp \cap U_1$, which is a complement to u in U_1. Since $\sigma\rho$ is an isometry, $\sigma\rho(u^\perp \cap U_1)$ is a complement to u in $\sigma(U_2)$. By induction, there is an isometry ϕ on $u^\perp$ that coincides with $\sigma\rho$ on $u^\perp \cap U_1$.

The linear map that fixes u and agrees with ϕ on $u^\perp$ is an isometry of V that agrees with $\sigma\rho$ on U_1. □

The following corollary is a form of *Witt cancellation*.

2.10.2 Corollary. *Let (V_1, q_1) and (V_2, q_2) be isometric quadratic spaces. Let U_1 and U_2 be subspaces of V_1 and V_2 respectively. If the radical of U_1 is zero and U_1 is isometric to U_2, then $U_1^\perp$ and $U_2^\perp$ are isometric.*

Proof. If σ is an isometry from V_2 to V_1, then $\sigma(U_2)$ is a subspace of V_1 isometric to U_1. So we may assume that $V_2 = V_1$ and $q_1 = q_2$. Now the theorem yields that there is an isometry σ such that $\sigma(U_1) = U_2$, and so $\sigma(U_1)^\perp = U_2^\perp$. As $\sigma(U_1)^\perp = \sigma(U_1^\perp)$, we are done. □

If the radical of (V, q) is zero, we say that q is *non-singular*. If 2 is invertible and $q(x) = x^T Ax$ for a symmetric matrix A, then q is non-singular is A is.

Suppose f_1 and f_2 are equivalent non-singular quadratic forms in the variables $x_1, \ldots, x_m$ and g_1 and g_2 are quadratic forms in a disjoint set of variables $y_1, \ldots, y_n$. Then if $f_1 + g_1$ and $f_2 + g_2$ are equivalent (as forms in $m+n$ variables), the previous corollary implies that g_1 and g_2 are equivalent. This is the form in which we will use Witt cancellation.

2.11 The Bruck-Ryser-Chowla Theorem

We need one preliminary result, due to Lagrange.

2.11.1 Theorem. *If n is a positive integer then $I_4 \approx nI_4$.*

Proof. Define the matrix

$$G := \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix}$$

and verify that

$$G^T G = (a^2 + b^2 + c^2 + d^2) I_4.$$

By a famous theorem due to Lagrange, every positive integer is equal to the sum of four squares and so the theorem follows. □

To place this result in some context, we observe that

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{pmatrix}$$

This implies that if $I_2 \approx nI_2$, then n must be the sum of two squares. Thus I_2 is not equivalent to $3I_2$. (It is not too hard to show that I_2 and nI_2 are equivalent if and only if n is the sum of two squares.)

The following necessary conditions for the existence of a symmetric design are due to Bruck, Rysler, and Chowla. We call $r - \lambda$ the *order* of a (v, k, λ)-design, and we denote it by n.

2.11.2 Theorem. *If there is a nontrivial symmetric (v, k, λ) design, one of the following two cases hold:*

(a) If v is even, $k - \lambda$ is a square.

(b) If v is odd, then the equation $(k-\lambda)x^2+(-1)^{\frac{v-1}{2}}\lambda y^2 = z^2$ has a nonzero integer solution.

Proof. First suppose v is even. Recall that

$$\widehat{N}\begin{pmatrix} I & 0 \\ 0 & -\lambda \end{pmatrix}\widehat{N}^{\mathrm{T}} = (k-\lambda)\begin{pmatrix} I & 0 \\ 0 & -\lambda \end{pmatrix}$$

Take determinants of both sides of this to get

$$(\det \widehat{N})^2(-\lambda) = (k-\lambda)^{(v+1)}(-\lambda)$$

From this we see that $(k-\lambda)^{(v+1)}$ is a square. This implies $(k-\lambda)$ is a square.

Now suppose that v is odd. There are two sub-cases to consider. If v is congruent to 1 modulo 4, we use Witt cancellation and the fact that $I_4 \approx (k-\lambda)I_4$ to cancel as many leading 4×4 blocks as possible, deducing as a result that

$$\begin{pmatrix} 1 & 0 \\ 0 & -\lambda \end{pmatrix} \approx \begin{pmatrix} n & 0 \\ 0 & -n\lambda \end{pmatrix}.$$

Since these forms are equivalent, they take the same integer values. Furthermore, since

$$\begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} n & 0 \\ 0 & -n\lambda \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = n$$

the right hand form takes the value n. Thus there are u and v such that

$$\begin{aligned} n &= \begin{pmatrix} u & v \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -\lambda \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} \\ &= \begin{pmatrix} u & v \end{pmatrix}\begin{pmatrix} u \\ -\lambda v \end{pmatrix} \\ &= u^2 - \lambda v^2 \end{aligned}$$

Here u and v are rational, and so by clearing denominators we get

$$nx^2 = y^2 - \lambda z^2, \quad x, y, z \in \mathbb{Z}.$$

This gives us the desired equation.

The second subcase is when v is congruent to 3 modulo 4. We have

$$x_1^2 + \cdots + x_v^2 - \lambda x_0^2 \approx ny_1^2 + \cdots + ny_v^2 - n\lambda y_0^2$$

and therefore

$$x_1^2 + \cdots + x_v^2 - \lambda x_0^2 + a_1^2 + na_2^2 \approx ny_1^2 + \cdots + ny_v^2 - n\lambda y_0^2 + b_1^2 + nb_2^2.$$

Consequently

$$(x_1^2 + \cdots + x_v^2 + a_1^2) - \lambda x_0^2 + na_2^2 \approx (ny_1^2 + \cdots + ny_v^2 + nb_2^2) - n\lambda y_0^2 + b_1^2$$

Since $v \equiv 3$ modulo 4, we can cancel the terms in parentheses, and deduce that

$$\begin{pmatrix} -\lambda & 0 \\ 0 & n \end{pmatrix} \approx \begin{pmatrix} 1 & 0 \\ 0 & -\lambda n \end{pmatrix}$$

Since

$$n = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} -\lambda & 0 \\ 0 & n \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

it follows that there are u and v such that

$$\begin{aligned} n &= \begin{pmatrix} u & v \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -\lambda n \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \\ &= \begin{pmatrix} u & v \end{pmatrix} \begin{pmatrix} u \\ -\lambda nv \end{pmatrix} \\ &= u^2 - \lambda nv^2. \end{aligned}$$

From this we see that

$$n + n\lambda v^2 = u^2$$

and multiplying both sides by n we have

$$n^2 + \lambda(nv)^2 = nu^2.$$

Setting $v_1 = nv$ we get

$$n^2 + \lambda v_1^2 = nu^2$$

and clearing denominators gives

$$n^2 z^2 = nx^2 - \lambda y^2.$$

Finally by defining $z_1^2 = n^2 z^2$ we have

$$z_1^2 = nx^2 - \lambda y^2.$$ □

2.12 Applications

We apply the Bruck-Ryser-Chowla conditions to projective planes. These are 2-designs with $\lambda = 1$ and $v = n^2 + n + 1$. There are two cases.

If $n \equiv 0, 3 \pmod 4$, then $v \equiv 1 \pmod 4$ and $nx^2 + y^2 = z^2$. Here $(x, y, z) = (0, 1, 1)$ is a non-zero integer solution.

If $n \equiv 1, 2 \pmod 4$, then $v \equiv 1 \pmod 4$ and $nx^2 = y^2 + z^2$. Thus we have

$$n = \left(\frac{y}{x}\right)^2 + \left(\frac{z}{x}\right)^2$$

which implies that $n = a^2 + b^2$ for some integers a and b.

2.12.1 Theorem. *If there is a projective plane of order n and $n \equiv 1, 2$ (mod 4), then n is the sum of two squares.* □

In particular, there is no projective plane of order six. (How many proofs of this can you find in the literature?)

Due to a difficult computation by Clement Lam from Concordia, we know that there is no projective plane of order 10, even though the conditions in the BRC Theorem are satisfied. However this is the only case we know where the BRC conditions for the existence of a symmetric design are satisfied, but the design does not exist.

Consider the problem of finding a non-zero solution to the equation

$$Ax^2 + By^2 + Cz^2 = 0$$

(where A, B, and C are integers). Assume that each pair from $\{A, B, C\}$ is coprime. Then necessary conditions for the existence of a non-zero solution are:

(a) A, B, C do not all have the same sign

(b) If the odd prime p divides A, then $-BC$ is a square modulo p.

(c) If the odd prime p divides B, then $-AC$ is a square modulo p.

(d) If the odd prime p divides C, then $-AB$ is a square modulo p.

For example if $p|A$, then

$$By^2 + Cz^2 = 0 \mod p$$

and therefore

$$B^2y^2 + BCz^2 = 0 \mod p,$$

from which it follows that $-BC$ must be a square modulo p. Legendre proved that the above four conditions are sufficient as well.

As examples, consider symmetric designs where $\lambda = 2$. Here

$$v = \frac{k(k-1)}{2} + 1 = \binom{k}{2} + 1.$$

The parameter sets for $7 \leq v \leq 79$ are as follows.

k	v	n
4	7	2
5	11	3
6	16	4
7	22	5
8	29	6
9	37	7
10	46	8
11	56	9
12	67	10
13	79	11

The designs with $k = 7$ and $k = 10$ do not exist because although v is even, n is not a square.

Consider $k = 8$ (so $(v, k, \lambda) = (29, 8, 2)$). Then the BRC equation is

$$6x^2 + 2y^2 = z^2$$

If there is a non-zero solution, then $2|z$. If $z = 2z_1$, then

$$6x^2 + 2y^2 - 4z_1^2 = 0$$

and so

$$3x^2 + y^2 - 2z_1^2 = 0.$$

Here $-BC = 2$, which is not a square modulo $A = 3$.

Chapter 3

Hadamard Matrices

A Hadamard matrix is an $n \times n$ matrix H with entries ± 1, such that

$$HH^T = nI.$$

For example

$$\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}.$$

We will meet many examples as we continue.

The are two families of operations we can apply that take a Hadamard matrix to a Hadamard matrix:

(a) Permute the rows and/or columns.

(b) Multiply all entries in a row or column by -1.

An arbitrary combination of these operations is a *monomial* operation. we say two Hadamard matrices are *monomially equivalent* if we can one from the other by monomial operations. (A Hadamard matrix need not be equivalent to its transpose.) A *monomial matrix* is a product of a permutation matrix and a diagonal matrix with diagonal entries equal to $\pm$; if M_1 and M_2 are monomial matrices and H is Hadamard, then $M_1 H M_2$ is Hadamard and is monomially equivalent to H.

A Hadamard matrix is *normalized* if all entries in its first row and column are equal to 1. Note that the equivalence class of H will contain many different normalized Hadamard matrices.

3.1 A Lower Bound

If $\mathcal{D}$ is a symmetric (v, k, λ) design, we define the difference $k - \lambda$ to be the *order* of $\mathcal{D}$, and we usually denote it by n. (This is consistent with the order of a finite projective plane.)

3.1.1 Theorem. *If $\mathcal{D}$ is a symmetric (v, k, λ) design, then*

$$4n - 1 \leq v \leq n^2 + n + 1.$$

Proof. Our concern here is with the lower bound, so we leave the upper bound as an exercise.

Let N be the incidence matrix of $\mathcal{D}$. Then each entry of $2N - J$ is equal to ± 1 and $(2N - J)\mathbf{1} = (2k - v)\mathbf{1}$. Consequently

$$\mathbf{1}^T(2N - J)^T(2N - J)\mathbf{1} = v(v - 2k)^2.$$

One the other hand

$$\begin{aligned}\mathbf{1}^T(2N - J)^T(2N - J)\mathbf{1} &= \mathbf{1}^T(4NN^T - 2NJ - 2JN^T + J^2)\mathbf{1} \\ &= \mathbf{1}^T(4nI + 4\lambda J - 2kJ - 2kJ + vJ)\mathbf{1} \\ &= \mathbf{1}^T(4nI - 4nJ + vJ) \\ &= 4nv + v^2(v - 4n).\end{aligned}$$

Consequently

$$v(v - 4n) + 4n = (v - 2k)^2 \geq 0$$

and therefore

$$v^2 - 4nv + 4n^2 + 4n \geq 4n^2,$$

which yields that

$$(v - 2n)^2 \geq 4n^2 - 4n.$$

If $n > 0$ then $n^2 - n$ is not square and thus we have slightly stronger inequality

$$(v - 2n)^2 \geq (2n - 1)^2.$$

This proves the lower bound. □

3.2 Equality

We decide what happens when a symmetric design has $v = 4n - 1$. Let $\overline{N}$ denote the matrix $2N - J$. Then

$$\overline{N}\overline{N}^T = 4nI - J$$

and therefore

$$\begin{pmatrix} \mathbf{1} & -\overline{N} \end{pmatrix} \begin{pmatrix} \mathbf{1} & -\overline{N} \end{pmatrix}^T = 4nI.$$

Since $\overline{N}\mathbf{1} = \mathbf{1}$, it follows that matrix

$$\begin{pmatrix} 1 & \mathbf{1}^T \\ \mathbf{1} & -\overline{N} \end{pmatrix}$$

is a normalized Hadamard matrix of order $4n$. Conversely, it is not too hard to show that a normalized Hadamard matrix of order $4n$ gives rise to a symmetric design with $v = 4n - 1$. We determine the parameters of this design in terms of n.

From the equation

$$v - 1 = \frac{k^2 - k}{\lambda}$$

we find that

$$(4n - 2)\lambda = (n + \lambda)(n + \lambda - 1) = n^2 + (2\lambda - 1)n + \lambda^2 - \lambda$$

and hence

$$0 = n^2 - (2\lambda + 1)n + \lambda(\lambda + 1) = (n - \lambda)(n - \lambda - 1).$$

If $\lambda = n$, then $k = 2n$. If $\lambda = n - 1$, then $k = 2n - 1$. Thus the parameters of the design are one of the pair

$$(4n - 1, 2n - 1, n - 1), \quad (4n - 1, 2n, n),$$

where the second pair is complementary to the first. A design with these parameters is called a *Hadamard design*. For a Hadamard matrix H, we get one Hadamard design for each possible way of normalizing H. In general these designs are not all isomorphic.

Hadamard matrices can also be used to construct a class of 3-designs. One approach is the assume that H has all entries in its first row equal to

one. Take the columns of H to be the points of the design. Each row of H other than the first determines a partition of the columns into two sets of size $n/2$, and we take these to be blocks. Thus the design we construct will have $2n-2$ blocks, and so it is a 2-design with parameters

$$\left(n, \frac{n}{2}, \frac{n}{2}-1\right).$$

Now

$$\frac{(2n-2)\binom{n/2}{3}}{\binom{n}{3}} = \frac{(2n-2)n(n-2)(n-4)}{8n(n-1)(n-2)} = \frac{n}{4}-1$$

and since this is an integer, our design could be a 3-design. We leave you to prove that it is.

3.3 The Kronecker Product

If A and B are matrices over the same field, we define their *Kronecker product* $A \otimes B$ to be the block matrix we get when we replace the ij-entry of A by $A_{i,j}B$, for all i and j. To give a very simple example, if

$$A = \begin{pmatrix} a \\ b \end{pmatrix}, \quad B = \begin{pmatrix} u & v \end{pmatrix},$$

then

$$A \otimes B = \begin{pmatrix} au & av \\ bu & bv \end{pmatrix}, \quad B \otimes A = \begin{pmatrix} au & av \\ bu & bv \end{pmatrix}.$$

We see that in general the Kronecker product is not commutative, but this is one of its few failings. It is bilinear, thus

$$A \otimes (xB + yC) = x(A \otimes B) + y(A \otimes C),$$
$$(xB + yC) \otimes A = x(B \otimes A) + y(C \otimes A).$$

One consequence of these identities is that

$$(xA) \otimes B = x(A \otimes B) = A \otimes (xB).$$

It is also easy to see that $(A \otimes B)^T = A^T \otimes B^T$.

The following provides one of the most important properties of the Kronecker product.

3.3.1 Lemma. *If the matrix products AC and BD are defined, then*

$$(A \otimes B)(C \otimes D) = AC \otimes BD. \qquad \square$$

It follows that

$$A \otimes B = (A \otimes I)(I \otimes B).$$

If x and y are vectors of the right orders, then we have the following useful special case of the above:

$$(A \otimes B)(x \otimes y) = Ax \otimes By.$$

If $e_1, \ldots, e_m$ is a basis for a vector space V over $\mathbb{F}$ and $f_1, \ldots, f_n$ is a basis for W, then the vectors

$$e_i \otimes f_j, \qquad 1 \leq i \leq m,\ 1 \leq j \leq n$$

form a basis for a vector space of dimension mn, which we denote by $V \otimes W$. We call $V \otimes W$ the *tensor product* of the vector spaces V and W. (Note that there will be elements of $V \otimes W$ that are not of the form $v \otimes w$, although the vectors of this form do span $V \otimes W$.

There is a unique linear mapping

$$P : V \otimes V \to V \otimes V$$

such that

$$P(x \otimes y) = y \otimes x$$

for all x and y. With respect to the basis

$$\{e_i \otimes e_j : 1 \leq i, j \leq \dim(V)\},$$

this is clearly a permutation matrix and $P^2 = I$, whence $P = P^T$. We call P the *flip operator* (or matrix).

3.3.2 Lemma. *If A and B are $m \times m$ matrices and P is the flip matrix, then $P(A \otimes B) = (B \otimes A)P$.*

Proof. We calculate:

$$P(A \otimes B)(x \otimes y) = P(Ax \otimes By) = By \otimes Ax = (B \otimes A)P(y \otimes x). \qquad \square$$

As a corollary it follows that if A is square then, although $A \otimes A^T$ need not be symmetric, the product $P(A \otimes A^T)$ is symmetric.

If A and B are two matrices of the same order, we define their *Schur product* $A \circ B$ by

$$(A \circ B)_{i,j} = A_{i,j}B_{i,j}, \qquad \forall i, j.$$

(It has been referred to as the "bad student's product".) This product is commutative and bilinear, and the matrix J is an identity for it. We find that

$$(A \otimes B) \circ (C \otimes D) = (A \circ C) \otimes (B \circ D).$$

3.4 Symmetric Hadamard Matrices

Suppose H is a Hadamard matrix of order n. Then H is a normal matrix and therefore there is a unitary matrix U and a diagonal matrix D such that

$$U^*HU = D.$$

Consequently

$$D = UHU^*$$

and since $n^{-1/2}H$ is unitary, it follows that $n^{-1/2}D$ is a product of of three unitary matrices. Therefore it is unitary and so its diagonal entries must have absolute value 1. We conclude that all eigenvalues of H have absolute value $\sqrt{n}$. (Note that a real normal matrix is symmetric if and only its eigenvalues are real.)

If you prefer a more elementary argument, suppose H is Hadamard of order n and z is an eigenvector for H with eigenvalue θ. Then

$$nz = H^THz = \theta H^Tz$$

and thus

$$nz^*z = \theta z^*H^Tz = \theta z^*H^*z = \theta(Hz)^*z = \theta\bar{\theta}z^*z.$$

Hence all eigenvalues of H have absolute value $\sqrt{n}$.

3.4.1 Lemma. *If H is a symmetric Hadamard matrix with constant diagonal of order n, then n is square.*

Proof. If all diagonal entries of H are equal to -1, we can multiply it by -1 and get a symmetric Hadamard matrix with 1's on the diagonal. Hence, we can assume the diagonal of H contains only 1's. From our remarks above, the eigenvalues of H are $\pm\sqrt{n}$. Assume $\sqrt{n}$ has multiplicity a.

As $\mathrm{tr}(H)$ is equal to the sum of its eigenvalues and as $\mathrm{tr}(H) = n$ we have

$$n = a\sqrt{n} - (n-a)\sqrt{n}.$$

If we divide this by $\sqrt{n}$, we see that $\sqrt{n} = 2a - n$, which is an integer. □

Examples

(a) $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is symmetric and its order is not a square.

(b) $\begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$ is symmetric with constant diagonal.

(c) If H_1 and H_2 are symmetric Hadamard matrices with constant diagonal, so is $H_1 \otimes H_2$.

(d) If H is Hadamard and P is the flip, then $P(H \otimes H^T)$ is symmetric. (What is its diagonal?)

3.5 Regular Hadamard Matrices

A Hadamard matrix is regular if its row and column sums are all equal. The class of regular Hadamard matrices is closed under Kronecker product.

3.5.1 Lemma. *If all rows of H have the same sum, then H is regular and its order is square.*

Proof. Suppose $H\mathbf{1} = k\mathbf{1}$ for some k. Then

$$kH^T\mathbf{1} = H^TH\mathbf{1} = n\mathbf{1}$$

and so

$$H^T\mathbf{1} = \frac{n}{k}\mathbf{1}.$$

This proves the regularity. Since all row sums are equal and all column sum are equal it follows that $\frac{n}{k} = k$ and $k = \pm\sqrt{n}$. □

If H is regular then the sum of the entries of H is $\pm n\sqrt{n}$. It can be shown that if K is an $n \times n$ Hadamard matrix then the absolute value of the sum of the entries of K is at most $n\sqrt{n}$, and equality holds if and only if the matrix is regular.

We can construct another interesting class of designs from regular Hadamard matrices.

3.5.2 Lemma. *Let H be $n \times n$ matrix with entries ± 1 and assume $N = \frac{1}{2}(J - H)$. Then H is a regular Hadamard matrix with row sum h if and only if N is the incidence matrix of a symmetric design with parameters $(4h^2, 2h^2 - h, h^2 - h)$.*

Design with these parameters, or their complements, are known as *Menon* designs.

3.5.3 Lemma. *A non-trivial symmetric (v, k, λ)-design is a Menon design if and only if $v = 4n$.*

In the exercises you will be offered the chance to prove that if $\mathcal{D}$ is a symmetric design on v and v is a power of 2, then $\mathcal{D}$ is a Menon design. Since the class of regular Hadamard matrices is closed under Kronecker product, we have an infinite supply of Menon designs.

3.6 Conference Matrices

An $n \times n$ matrix C is a conference matrix if $c_{i,i} = 0$ for all i and $c_{i,j} = \pm 1$ if $i \neq j$ and $CC^T = (n-1)I$.

Examples:

$$\begin{pmatrix} 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

A conference matrix is normalized if all non-zero entries in the first row are equal and all non-zero entries in the first column are equal.

If $n \geq 2$ and there is a conference matrix of order n, then n is even. We can say more.

3.6.1 Theorem. *If C is a conference matrix of order n and $n \equiv 2 \pmod 4$ then $n - 1$ is the sum of two squares.*

Proof. We have

$$CC^T = (n-1)I$$

and so the symmetric bilinear forms I_n and $(n-1)I_n$ are equivalent. By Witt cancellation we deduce that

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \approx \begin{pmatrix} n-1 & 0 \\ 0 & n-1 \end{pmatrix}.$$

Hence there is an invertible matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with

$$\begin{pmatrix} n-1 & 0 \\ 0 & n-1 \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2+b^2 & . \\ . & . \end{pmatrix}.$$

which implies $a^2 + b^2 = n - 1$. □

3.6.2 Lemma. *If C is a skew symmetric conference matrix then $I + C$ is a Hadamard matrix.*

Proof.

$$(I+C)(I+C)^T = (I+C)(I-C) = I - C^2 = I + CC^T = nI$$

□

3.6.3 Lemma. *If C is a symmetric conference matrix, then*

$$\begin{pmatrix} C+I & C-I \\ I-C & C+I \end{pmatrix}$$

is a Hadamard matrix.

Proof. Compute HH^T. □

3.6.4 Theorem. *Let C be a conference matrix of order $n \times n$ with all entries in its first row and column non-negative, and let C_1 be the matrix we get by deleting the first row and column of C. Then n is even and if $n \equiv 2$ modulo 4, then C_1 is skew symmetric; if $n \equiv 0$ modulo 4 then C_1 is symmetric.*

Proof. Let c_1, c_2 and c_3 denote the first three rows of C. We suppose

$$x := C_{2,3}, \quad y := C_{3,2}.$$

Then

$$(c_1 + c_2) \circ (c_1 + c_3) = c_1 \circ c_1 + c_2 \circ c_1 + c_1 \circ c_3 + c_2 \circ c_3.$$

The inner product of the left side with $\mathbf{1}$ has the form $3 + x + y + 4m$; since the rows c_1, c_2 and c_3 are orthogonal, the inner product of the right side with $\mathbf{1}$ is $n - 1$. Therefore

$$n - 1 = 3 + x + y \mod 4.$$

Since $x + y$ is even it follows that n is even and that $n \equiv x + y$ modulo 4. □

If C is a normalized conference matrix and C_1 is the matrix we get by deleting the first row and column from C, then

$$(n-1)I = \begin{pmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & C_1 \end{pmatrix} \begin{pmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & C_1^T \end{pmatrix} = \begin{pmatrix} n-1 & \mathbf{1}^T C_1^T \\ C_1\mathbf{1} & C_1 C_1^T + J \end{pmatrix}$$

and consequently

$$C_1\mathbf{1} = 0, \quad C_1 C_1^T = (n-1)I + J.$$

3.7 Paley Matrices

We now start on the construct of conference matrices of order $q + 1$, one for each odd prime power q. These matrices will be skew symmetric when $q \equiv 3 \mod 4$ and symmetric if $q \equiv 1 \mod 4$. In the former case we will then get Hadamard matrices of order $2q+2$. (Using these and the Kronecker product, we can construct Hadamard matrices of all orders $4m$ between 8 and 96, except the case $4m = 92$.)

Let q be an odd prime power and let $\mathbb{F}$ be a finite field of order q. (All finite fields of the same size are isomorphic). If $a \in \mathbb{F}$ then

$$\mathcal{X}(a) = \begin{cases} 1 & a \text{ is a square, but not } 0 \\ -1 & a \text{ is not a square} \\ 0 & a = 0 \end{cases}$$

We call $\mathcal{X}$ the *Legendre function* or *quadratic character*.

The Paley matrix is the $q \times q$ matrix with rows and columns indexed by $\mathbb{F}$, such that $M_{a,b} = \mathcal{X}(b-a)$ for $a, b \in \mathbb{F}$.

Note that all the diagonal entries of M are zero. The off-diagonal entries are ± 1. If $q \equiv 1 \mod 4$, then M is symmetric. If $q \equiv 3 \mod 4$, M is skew-symmetric ($M^T = -M$).

If C is a normalized conference matrix, then the matrix we get by deleting the first row and column is its *core*. A Paley matrix is the core of a conference matrix. (The proof is an exercise).

Chapter 4

Projective Planes

4.1 Projective Planes

A *projective plane* is a thick incidence structure such that each pair of points lies on exactly one block and each pair of blocks has exactly one point incident with both of them. In this context blocks are always called lines, and a block is usually identified with the the set of points incidence with it.

We offer a construction. Let V be a vector space of dimension three over a field $\mathbb{F}$, for example $\mathbb{Z}_p$. Let $\mathcal{P}$ denote the incidence structure with the 1-dimensional subspaces of V as its points and the 2-dimensional subspaces as its lines, where a point is incident with a line if the corresponding 1-dimensional subspace is contained in the 2-dimensional subspace. We will be chiefly concerned with the case where $\mathbb{F}$ is finite.

If $|\mathbb{F}| = q$, then $|V| = q^3$. Each 1-dimensional subspace contains $q - 1$ non-zero vectors, and these sets of non-zero vectors form a partition of the $q^3 - 1$ non-zero vectors in V. Hence $\mathcal{P}$ has

$$\frac{q^3 - 1}{q - 1} = q^2 + q + 1$$

points. If a is a non-zero vector in V, then $\ker(a^T)$ is a 2-dimensional subspace of V and $\ker(a^T) = \ker(b^T)$ if and only if a and b are non-zero scalar multiples of each other. It follows that there are exactly $q^2 + q + 1$ lines in $\mathcal{P}$. Each line contains exactly $q + 1$ points, and a simple counting argument yields that each point is on exactly $q + 1$ lines. Thus $\mathcal{P}$ is point and block regular.

4.1.1 Lemma. *A finite projective plane is a symmetric 2-*$(n^2+n+1, n+1, 1)$ *design for some integer.*

Proof. □

The *order* of a projective plane is one less than the number of points in a line. Thus the plane constructed above from the field of order q has $q+1$ points on each line, and so the order of the plane is q.

4.1.2 Theorem. *Let G be the incidence graph of an incidence structure $\mathcal{P}$. Then $\mathcal{P}$ is a projective plane if and only if G has minimum valency at least three, diameter three and girth six.*

Proof. □

4.2 Near Misses

Suppose $\mathcal{P}$ is a projective plane and α is a collineation of $\mathcal{P}$. Let $\mathcal{F}$ denote the incidence structure that consists of the points and lines that are fixed by α.

If x and y are distinct points fixed by α and a is the line that contains x and y, then a^α is a line that contains x^α and y^α. But $x^\alpha = x$ and $y^\alpha = y$, and consequently $a^\alpha = a$. Similarly we see that if α fixes two distinct lines a and b, it must fix their point of intersection. We conclude that $\mathcal{F}$ is both a linear space and a dual linear space. If $\mathcal{F}$ is thick, it must be a projective plane. Note that the incidence structure consisting a line and all the points on it is a linear space and a dual linear space. What other possibilities can arise?

4.2.1 Theorem. *If $\mathcal{F}$ is a linear space and a dual linear space, but not a projective plane then $\mathcal{F}$ is isomorphic to one of the following:*

(a) A set of points and a single line that contains all the points.

(b) A set of lines incident with a single point.

(c) A line ℓ with at least two points on it, a point x not on the line and, for each point y on ℓ, a further line incident with x and y (but not with any other points).

Proof. □

Note that the first cases form a dual pair and that the third is self-dual. An *arc* in an incidence structure is a set of points such that no three are collinear; a *k-arc* is an arc of size k.

4.2.2 Theorem. *Suppose $\mathcal{P}$ is a linear space and a dual linear space. Then $\mathcal{P}$ is projective plane if and only if it contains a 4-arc.*

Proof. □

4.3 De Bruijn and Erdős

An *antiflag* in an incidence structure is a pair (x, β) where β is a line and x is a point not on β. (If x is on β, we have a *flag*.)

4.3.1 Theorem. *Let $\mathcal{D}$ be a linear space on v points. Then $\mathcal{D}$ has either one line or at least v lines; if it has v lines, then it is a dual linear space.*

Proof. If x is a point of $\mathcal{D}$, let $r(x)$ denote the number of lines on x and if β is a line, let $k(\beta)$ denote the number of points incident with the line.

Assume that there is more than one line. If the point x is not on the line β, then there are $k(\beta)$ lines that join x to points on β, so $r(x) \geq k(\beta)$.

Let b be the number of lines in $\mathcal{D}$ and suppose $b \leq v$. Then $vr(x) \geq bk(\beta)$ and so

$$b(v - k(\beta)) \geq v(b - r(x)). \tag{4.3.1}$$

Now using (4.3.1),

$$1 = \sum_{x,\beta : x \notin \beta} \frac{1}{v(b - r(x))} \geq \sum_{x,\beta : x \notin \beta} \frac{1}{b(v - k(\beta))} = 1.$$

This implies that all the inequalities in (4.3.1) are equalities. Hence $v = b$ and $r(x) = k(\beta)$ for each antiflag (x, β). The latter implies that each line on x contains a point of β, and hence that any two lines intersect. □

The above proof is due to John Conway. Cameron derives the theorem of De Bruijn and Erdős from Hall's condition for the existence of matchings.

4.4 Ovals and Hyperovals

An *arc* in an incidence structure is a set of points such that no three are incident with the same block and a k-arc is an arc of size k. Here we are concerned with arcs in projective planes.

4.4.1 Lemma. *Let $\mathcal{P}$ be a projective plane of order n, and let α_i denote the number of ordered sets of i points that form an i-arc. Then:*

(a) $\alpha_1 = n^2 + n + 1$.

(b) $\alpha_2/\alpha_1 = (n^2 + n)$.

(c) $\alpha_3/\alpha_2 = n^2$.

(d) $\alpha_4/\alpha_3 = (n-1)^2$.

Proof. □

It follows from this result that a projective plane always contains a 4-arc. How large can an arc be?

4.4.2 Lemma. *Let $\mathcal{P}$ be a plane of order n. Then any arc in $\mathcal{P}$ has order at most $n+2$, if n is odd then an arc has order at most $n+1$.*

Proof. Let C be an arc in a plane of order n, and let x be a point in C. For each point y in $C \setminus x$ there is a line through x; these lines are distinct since no line contains three points from C. Since there are exactly $n+1$ lines on x,

$$|C \setminus x| \leq n + 1$$

and therefore $|C| \leq n + 2$.

Now suppose that C is an arc of size $n + 2$. If $x \in C$, then by the first paragraph each line on x must meet C in a second point. So each line of the plane that meets C must contain two points from C. Suppose y is a point not in C. Then the lines through y that meet C each meet it in two points. Thus they partition C into distinct pairs and consequently $|C|$ is even. Therefore n is even. □

An *oval* in a projective plane of order n is an $(n+1)$-arc. A *hyperoval* is an $(n+2)$-arc.

4.5 Affine Planes

An incidence structure $\mathcal{A}$ is an affine plane if

(a) Each two distinct points lie on exactly one line.

(b) If ℓ is a line and x is a point not on ℓ, there is a unique line through x and disjoint from ℓ.

(c) There are three points not all one one line.

In (b), we say that the line on x is *parallel* to ℓ.

The vector and the cosets of the 1-dimensional subspaces of a 2-dimensional vector space V over $\mathbb{F}$ form an affine plane.

We can also construct affine planes from projective planes. Suppose $\mathcal{D}$ is an incidence structure with point set V. Let B be a block in $\mathcal{D}$. We form a new incidence structure with point-set $V \setminus B$; its blocks are the blocks of $\mathcal{D}$ distinct from B and a block is incident with a point in the new structure if it was incident with it in $\mathcal{D}$. We call the new structure the *residual* of $\mathcal{D}$ relative to B.

4.5.1 Theorem. *If $\mathcal{P}$ is a projective plane and ℓ is a line in it, the residual of $\mathcal{P}$ relative to ℓ is an affine plane.*

Proof. Let $\mathcal{A}$ denote the residual structure. It is a partial linear space because $\mathcal{P}$ is.

We find a 3-arc. Let x be a point of $\mathcal{P}$ not on ℓ. Choose two lines on x and, on both of these choose a point distinct from x and not on ℓ. If these points are y and z, then x, y and z are not collinear.

Now let m be a line of $\mathcal{P}$ not equal to ℓ and let x be a point not on m (and not on ℓ). Let z be the intersection of ℓ and m. Then in $\mathcal{A}$, the unique line joining x and z is disjoint from ℓ. So there is a line through x parallel to ℓ and we have to show it is unique. But a second parallel would be a second line in $\mathcal{P}$ that contains x and z. Since this is impossible, $\mathcal{A}$ is an affine plane.

4.5.2 Lemma. *Parallelism is an equivalence relation on the lines of an affine plane.*

Proof. Let ℓ, m and n be lines in the affine plane $\mathcal{A}$ such that m is disjoint from ℓ and n is disjoint from m. Assume by way of contradiction that ℓ and

n meet in the point x. Then we have found two lines on x that are disjoint from n, which contradicts our axioms. □

The equivalence classes with respect to parallelism are called *parallel classes*.

You may check that if $\mathcal{P}$ is a plane of order n, then the affine plane we obtain as the residual relative to a line ℓ is a 2-$(n^2, n, 1)$ design.

4.5.3 Theorem. *A 2-$(n^2, n, 1)$ design where $n \geq 2$ is an affine plane.*

Proof. Let $\mathcal{D}$ be a 2-$(n^2, n, 1)$ design. Since $\lambda = 1$, it is a partial linear space. If x is a point not in the block β then x together with any two points from β is a 3-arc.

We see that $r = n + 1$. If x is a point not on the block β, there are exactly n lines that join x to points in β and therefore there is a unique block on x disjoint from ℓ. □

The process of going from a projective plane to an affine plane is reversible:

4.5.4 Theorem. *Let $\mathcal{A}$ be an affine plane and let $\mathcal{E}$ be the set of parallel classes of $\mathcal{A}$. Let $\mathcal{P}$ be the incidence structure defined as follows.*

(a) The points of $\mathcal{P}$ are the points of $\mathcal{A}$ and the elements of $\mathcal{E}$.

(b) The lines of $\mathcal{P}$ are the lines of $\mathcal{A}$, and one new line ℓ_∞.

(c) The line ℓ_∞ is incident only with the elements of $\mathcal{E}$.

(d) A line of $\mathcal{A}$ is incident with the element of $\mathcal{E}$ that contains it.

(e) The incidence between points and lines of $\mathcal{A}$ are unchanged.

Then $\mathcal{P}$ is a projective plane and $\mathcal{A}$ is its residual relative to ℓ_∞. □

4.6 Affine Planes from Spreads

Suppose G is a group of order n and let $\mathcal{H} = \{H_1, \ldots, H_r\}$ be a set of subgroups of G of size m such that the sets $H_i \setminus 1$ partition $G \setminus 1$. Then, as you're invited to show in the exercises, the incidence structure with points set G and with the cosets of the subgroups in $\mathcal{H}$ as lines is a resolvable

2-design. We can use this idea to construct affine planes, and the planes we get in this way are not usually Desarguesian.

A group G will be the vector space $V(2d, q)$ over a field of order q. The subgroups H_i will be subspaces of dimension d, hence

$$|\mathcal{H}| = \frac{q^{2d} - 1}{q^d + 1} = q^d + 1.$$

If $|\mathcal{H}| = q^d + 1$, we call it a *spread*.

Two subspaces that intersect in the 0-subspace are called *skew*. We can present our subspaces as the column spaces of $2d \times d$ matrices. Before we do this, we normalise things somewhat—note if (H_1, H_2) and (H_3, H_4) are two pairs of skew d-dimensional subspaces, there is an invertible linear map on V that sends the first pair to the second. If $V(\infty)$ and $V(0)$ respectively are the column spaces of the matrices

$$\begin{pmatrix} 0 \\ I \end{pmatrix}, \qquad \begin{pmatrix} I \\ 0 \end{pmatrix},$$

we may assume that they are elements of $\mathcal{H}$. The next step is to note that the the column space of the $2d \times d$ matrices

$$\begin{pmatrix} A \\ B \end{pmatrix}, \qquad \begin{pmatrix} C \\ D \end{pmatrix}$$

are skew if and only the matrix

$$\begin{pmatrix} A & C \\ B & D \end{pmatrix}$$

is invertible. Hence the column space of

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

is skew to $V(\infty)$ if and only if A is invertible. If A is invertible, then the column spaces of

$$\begin{pmatrix} A \\ B \end{pmatrix}, \qquad \begin{pmatrix} I \\ BA^{-1} \end{pmatrix}$$

are equal. Finally the column space $V(M)$ of

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

is skew to $V(\infty)$ and it is skew to $V(0)$ if and only if M is invertible.

Hence if a spread $\mathcal{H}$ exists we may assume it contains $V(\infty)$, $V(0)$ and matrices $V(M)$, but we need to determine the conditions on the matrices M. For this, note that $V(M)$ and $V(N)$ are skew if and only if

$$\begin{pmatrix} I & I \\ M & N \end{pmatrix}$$

is invertible; as the determinant of this matrix is $N - M$ it follows that a necessary condition for the subspaces

$$V(\infty),\ V(0),\ V(M_1), \ldots, V(M_{q^d-1})$$

to form a spread is that the matrices M_i are invertible and $M_i - M_j$ is invertible if $i \neq j$. If we set $M_0 = 0$, our requirement is that $M_i - M_j$ is invertible whenever $0 \leq i, j \leq q^d - 1$ and $i \neq j$. (Such a set of matrices is called a *spread set*).

Chapter 5

Groups and Matrices

5.1 Difference Sets

Let Γ be an abelian group and let C be a subset of Γ. If $g \in \Gamma$ we define the set $C + g$ by

$$C + g := \{c + g : c \in C\}.$$

We call it the *translate* of C under g. We obtain an incidence structure by taking the elements of Γ as points and the translates of C as the blocks. So there are $|\Gamma|$ points and at most $|\Gamma|$ blocks.

The canonical example arises when $\Gamma = \mathbb{Z}_7$ and $C = \{0, 1, 3\}$. The corresponding incidence structure is a projective plane of order two.

5.1.1 Theorem. *Let C be a subset of the abelian group Γ with $|C| \geq 3$ and suppose that for each non-zero element g of G there is a unique ordered pair (x, y) of elements of C such that $g = y - x$. Then the incidence structure with the elements of Γ as its points and the translates of C as its lines is a projective plane.*

Proof. Suppose x and y are distinct elements of C and u and v are elements of C and for some element h of C,

$$u = x + h, \quad v = y + h.$$

Then

$$v - u = y - x$$

and so if the stated condition on C holds, then any distinct two translates of C have at most one element in common. We also see that if $v - u = y - x$,

then the translate $C + h$ of C contains u and v. It follows that each pair of distinct elements of C lies in a translate of C, and therefore our incidence structure is a linear space.

Since the number of translates of C is at most $|\Gamma|$, it follows from De Bruijn and Erdős that our incidence structure is a dual linear space. Since $|C| \geq 3$ it is thick, and therefore it is a projective plane. □

It is not strictly necessary to appeal to De Bruijn and Erdős in the above proof, but it is easier.

A set C satisfying the condition of the theorem is an example of a *planar difference set*. There are two useful generalizations. There is the case where the set of differences covers each non-zero element of G exactly λ times, for some positive integer λ. And there is the case where we take several sets $C_1, \ldots, C_k$ all of the same size, such that each non-zero element of G occurs exactly once as a difference of two elements from the same set C_i. (And there is a common generalization of these two cases; all these generalizations are known as *difference sets*.)

As an exercise, produce planar difference sets in $\mathbb{Z}_{13}$ and $\mathbb{Z}_{21}$.

5.2 Ovals and Difference Sets

If D is a subset of an abelian group G define

$$-D := \{-g : g \in D\}.$$

5.2.1 Lemma. *If D is a planar difference set in an abelian group G, then $-D$ is an oval.*

Proof. Let $\mathcal{P}$ be the plane generated by D. If S is a set of points in $\mathcal{P}$, then S is an arc if and only if, for each translate $D + g$ of D,

$$|S \cap (D + g)| \leq 2.$$

Now if $-a$, $-b$ and $-c$ are elements of $-D$ and

$$-a = u + g,\ -b = v + g,\ -c = w + g$$

where $u, v, w \in D$ and $g \in G$, then

$$-g = a + u = b + v$$

and so

$$a - b = v - u.$$

Since D is a planar difference set, this implies that $a = v$ and $b = u$. Since

$$-g = b + v = c + w,$$

we also find that $b = w$ and $c = v$. We conclude that $-D$ is an oval in $\mathcal{P}$. □

Let $\mathbb{F}$ be a field of order q. You may show that the vector

$$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

together with the vectors

$$\begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix}$$

for $t \in \mathbb{F}$ forms an oval. If $a \in \mathbb{F}^3$, then $\ker(a^T)$ is a line; this line contains the point with coordinate vector x if $a^T x = 0$. You should use this to show that each line contains at most two points from the oval just given. The oval just given is an example of a *conic*.

5.3 Group Matrices

Let G be a finite group. We say that a matrix M is a *group matrix* or a *G-matrix* if its rows and columns are indexed by the elements of G, and there is a function ψ in G such that

$$M_{g,h} := \psi(hg^{-1}).$$

Normally the first row and column of a group matrix will be indexed by the identity element and ψ will take values in a field. A group matrix over $\mathbb{Z}_n$ is better known as a *circulant*.A circulant is a matrix where each row is the cyclic right shift of the row 'above' it.

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{pmatrix}$$

Thus a circulant is a group matrix over $\mathbb{Z}_n$.

If M is a G-matrix then so is M^T (exercise). The set of all group matrices over G with values in the ring R will be denoted by $R[G]$. A *difference set* over G is a G-matrix that is the incidence matrix of a symmetric design.

5.3.1 Lemma. *If M and N are group matrices, then so are MN and $M \circ N$.*

It follows that the set of all group matrices over G with values in $\mathbb{F}$ is a matrix algebra—a vector space of matrices that contains I and is closed under matrix multiplication. It is also closed under transpose and under field automorphisms (for example, under complex conjugation when $\mathbb{F} = \mathbb{C}$.) If we need to be precise, it is a faithful representation of the group algebra of G over $\mathbb{F}$ (but fortunately we do not need to be precise).

A function μ on G is constant on conjugacy classes if

$$\mu(g^{-1}ag) = \mu(a)$$

for all g in G. (Suppose for example that the elements of G were matrices and μ is the trace function.) We say a group matrix is *central* if its defining function is constant on conjugacy classes.

5.3.2 Lemma. *Let M be a group matrix based on the function μ. Then M commutes with all G-matrices if and only if μ is constant on the conjugacy classes of G.*

There is a natural basis for the space of G-invariant matrices. Define P_x, for $x \in G$ by

$$(P_x)_{g,h} = \begin{cases} 1, & gh^{-1} = x \\ 0, & \text{otherwise} \end{cases}$$

We can check that $P_x P_y = P_{xy}$ (exercise). Note that P_x is a permutation matrix. If e_g is the standard basis for $\mathbb{F}^G$ (indexed by the elements of G), then

$$P_g e_x = e_{xg^{-1}}.$$

If ψ and φ are functions on G, their *convolution* $\psi * \varphi$ is given by

$$(\psi * \varphi)(g) := \sum_{x \in G} \psi(x)\varphi(x^{-1}g)$$

If M is a G matrix and N is an H-matrix, then $M \otimes N$ is a group matrix for $G \times H$.

5.4 Eigenvalues and Eigenvectors

We consider the space of group matrices over a finite abelian group G. If η is a function on G, then we define $M(\eta)$ to be the matrix with columns

$$P_g\eta, \quad g \in G.$$

We have

$$M(\eta) = \sum \eta(g) P_g^{-1}.$$

The matrices P_g for g in G are normal and commute, hence they have a common orthonormal basis of eigenvectors. If η is a common eigenvector, there is a complex-valued function λ_g on G such that

$$P_g\eta = \lambda_g\eta.$$

This function is a homomorphism from G to the group of complex numbers of norm 1; such a function is a *character* of G. Clearly each eigenvector determines a character of G. Remarkably, the converse is also true:

$$P_g \sum_x \lambda_x e_x = \sum_x \lambda_x e_{xg^{-1}} = \sum_y \lambda_{yg} e_y = \lambda_g \sum_y \lambda_y e_y.$$

If λ and μ are two characters of G, then the map

$$g \mapsto \lambda_g\mu_g$$

is again a homomorphism from G into the $\mathbb{C}^*$, and so it is a character. Hence the set of characters of G forms a group, called the *dual group* of G and denoted by G^*. It is isomorphic to G. (Prove this for cyclic groups, then show that $(G \times H)^* = G^* \times H^*$.)

Fix an isomorphism from G to G^*, and let χ_g be the character assigned to g in G. If M is a G-matrix, then we define $\Phi(M)$ to be the matrix with rows and columns indexed by G, such that $(\Phi(M))_{g,h}$ is equal to the eigenvalue of M on the character $\chi_{hg^{-1}}$. Thus $\Phi(M)$ is a G-matrix,

$$\Phi(I) = J$$

and, for any two group matrices M and N,

$$\Phi(MN) = \Phi(M) \circ \Phi(N).$$

We also have

$$\Phi(J) = nI.$$

5.4.1 Theorem. *We have* $\Phi(M \circ N) = n^{-1}\Phi(M)\Phi(N)$ *and* $\Phi^2(M) = nM^T$.

5.5 Multipliers

A *multiplier* of a difference set in G is an automorphism of G that is also an automorphism of $\mathcal{D}$.

5.5.1 Theorem. *Suppose $\mathcal{D}$ is a symmetric (v, k, λ) design and $p \mid n$ and $p > \lambda$. If $S \subseteq V$ such that $|S| = k$ and its characteristic function $x_S \in \mathrm{col}_p(\mathcal{D})$, then S is a block.*

Proof. Since $p > \lambda$, we see that p does not divide k. Let $\beta_1, \ldots, \beta_v$ denote the blocks of $\mathcal{D}$. Since $x_S \in \mathrm{col}_p(\mathcal{D})$, we can write

$$x_S = \sum_i a_i x_{\beta_i}, \quad a_i \in GF(p).$$

Then

$$k = \mathbf{1}^T x_S = \sum_i a_i \mathbf{1}^T x_{\beta_i} = k \sum a_i \tag{5.5.1}$$

If β is some block in $\mathcal{D}$ we also have

$$\lambda = x_\beta^T x_S = \sum a_i x_\beta^T x_{\beta_i} = k \sum a_i = k \pmod p$$

and, since $p > \lambda$, this implies that $|S \cap \beta| \geq \lambda$.

We now prove that if S is a k-subset of V and $|S \cap \beta| \geq \lambda$ for all blocks β, then S is a block. We have

$$NN^T = nI + \lambda J$$

and consequently

$$N\left(N - \frac{\lambda}{k}J\right)^T = nI.$$

If $|S \cap \beta| \geq \lambda$ for all blocks λ then $x_S^T N \geq \lambda \mathbf{1}^T$ and $x_S^T \left(N - \frac{k}{\lambda}J\right) \geq \mathbf{0}$. We have

$$x_\beta^T \left(N - \frac{\lambda}{k}J\right) = ne_\beta$$

Let $a^T = x_S^T \left(N - \frac{\lambda}{k}J\right)$. Then,

$$\begin{aligned} a^T &= \sum a_\beta e_\beta^T \\ &= \frac{1}{n} \sum a_\beta n e_\beta^T \\ &= \frac{1}{n} \sum a_\beta x_\beta^T \left(N - \frac{\lambda}{k}J\right) \end{aligned}$$

and therefore

$$s_S^T \left(N - \frac{\lambda}{k} J \right) = \sum \frac{a_\beta}{n} x_\beta^T \left(N - \frac{\lambda}{k} J \right).$$

Since $N - \frac{k}{\lambda} J$ is invertible, we find that

$$x_S = \frac{1}{n} \sum a_\beta x_\beta$$

where $a_\beta \geq c$ for all β. Since x_S has weight k, only a_β can be non-zero. So $x_S = x_\beta$ for some block β. □

There is a useful way to view the second part of the above proof. If M is a matrix, then the vectors y such that $y^T M \geq 0$ form a *convex cone*, that is, a set of vectors in a real vector space closed under addition and multiplication by non-negative scalars. The set of all non-negative linear combinations of the columns of M is also a convex cone, dual to the first. In the above proof we are arguing that the rows of N generate the dual to the cone generated by the rows of $N - \frac{la}{k} J$.

5.5.2 Theorem. *Let $\mathcal{D}$ be a symmetric design given by a difference set in the abelian group G. If $p \mid n$ and $p > \lambda$, then p is a multiplier of the difference set.*

Proof. We can assume $N \in \mathbb{F}_p(G)$; N is a sum of permutation matrices from G with coefficients in $\mathbb{F}_p$.

We can write N as $\sum c_i P_i$ where $c_i = 0, 1$, and then

$$N^p = \sum c_i^p P_i^p = \sum c_i P_i^p.$$

If $P \in G$ then $PN - NP$, as $\mathrm{col}(N)$ is invariant under each element of G. Also,

$$N^p = N^{p-1} N$$

whence we see that $\mathrm{col}(N^p) \subseteq \mathrm{col}(N)$. Each column of N^p is the characteristic vector of a subset of size k, hence it must be the characteristic vector of a block of $\mathcal{D}$. □

5.5.3 Lemma. *A multiplier of a symmetric design over a group G fixes at least one block.*

Proof. We can represent the action of the multiplier on points by the permutation matrix P. Then each column of PN is a block, and $PN = NQ$ for some permutation matrix Q. Since N^{-1} exists,

$$Q = N^{-1}PN$$

and therefore $\mathrm{tr}(Q) = \mathrm{tr}(P)$. Therefore the number of blocks fixed by Q equals the number of points fixed by P. Since P fixes 0, we are done. □

5.5.4 Theorem. *If $(v, k) = 1$ and $\mathcal{D}$ is given by a difference set in an abelian group, there is a block which is fixed by every multiplier.*

Proof. Let β be a block and let h be the product of its elements. If $g \in G$, then the product of the elements in the block βg is $g^k h$. Since k and n are coprime, the map $g \mapsto g^k$ is an automorphism of G and therefore there is a unique element g of G such that $g^k h = 1$. This shows that there is a unique block γ such that the product of its elements if 1. Clearly this block is fixed by all multipliers. □

Chapter 6

Orthogonal Arrays

6.1 Latin Squares

Traditionally a *Latin square of order* n is an $n \times n$ array with entries from $\{1, \ldots, n\}$ such that each integer occurs exactly once in each row and column. If we allow ourselves to use any set of size n for the entries, we see that the multiplication table of a group is a Latin square. Therefore there is a Latin square of order n for each non-negative integer n.

There is an alternative definition which works better. Suppose A is matrix that represents a latin square of order n. Then we can have n^2 triples

$$(i, j, A_{i,j}).$$

We can now write down an $n^2 \times 3$ matrix with these triple as rows. (Although for typographical reasons we might write down the transpose instead.) This matrix has the property that each ordered pair of columns contains each ordered pair from $\{1, \ldots, n\}$ exactly once. Conversely any $n^3 \times 3$ matrix over $\{1, \ldots, n\}$ with this property comes from a latin square.

We can now introduce a key concept: an *orthogonal array* $OA(n, k)$ over $\{1, \ldots, n\}$ is a matrix with k columns and entries from $\{1, \ldots, n\}$, such that each ordered pair of columns contains each ordered pair of elements from $\{1, \ldots, n\}$ exactly once. It follows that an orthogonal array has exactly n^2 rows. An orthogonal array $OA(n, 2)$ is more or less the edge set of the complete bipartite graph $K_{n,n}$. An $OA(n, 3)$ is a latin square. Two orthogonal arrays are *equivalent* if we can get one from the other by a combination. of permuting rows, columns or symbols.

We can generalize the concept of orthogonal array to that of an orthogonal array with index λ, where each ordered pair occurs exactly λ times in each pair of columns. An orthogonal array has strength at least t if, for each s with $1 \leq s \leq t$, each ordered s-tuple occurs exactly λ_s times in each set of s columns.

We can describe orthogonal arrays as incidence structures. Assume we have an $OA(n, k)$. The blocks of the structure are the rows of the array. We define the points as ordered pairs (i, α) where $i \in \{1, \ldots, k\}$ and $\alpha \in \{1, \ldots, n\}$; a point (i, α) is incident with the blocks with i-th coordinate equal to α. We have $3n$ points and v^2 blocks. It is not hard to show that this structure is a partial linear space: any two point with different first coordinate lie in a unique block, while no block is incident with two distinct points that have distinct first coordinate.

The dual structure is resolvable. The n points with first coordinate i form a parallel class and this gives k parallel classes whose union is the point set.

6.2 Examples

Let $\mathbb{F}$ be a finite field of order q. If $a_1, \ldots, a_{q-1}$ are the non-zero elements of $\mathbb{F}$, form the array with q^2 rows of the form

$$(x, y, x + a_1 y, \ldots, x + a_{q-1} y)$$

This is an orthogonal array. If $|\mathbb{F}| = 3$, the rows are

$$\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 2 & 2 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 2 & 0 \\ 1 & 2 & 0 & 2 \\ 2 & 0 & 2 & 2 \\ 2 & 1 & 0 & 1 \\ 2 & 2 & 1 & 1 \end{array}$$

Suppose $\mathcal{A}$ and $\mathcal{B}$ are orthogonal arrays with k columns with elements from M and N respectively. If (i, j, α) and (k, ℓ, β) are rows of $\mathcal{A}$ and $\mathcal{B}$

respectively, define their product to be

$$((i,k),(j,\ell),(\alpha_1,\beta_1),\ldots,(\alpha_{k-2},\beta_{k-2}))$$

The set of all products forms an array over $M \times N$, and you may verify that this is an orthogonal array.

We present an application. Suppose we have a 2-$(v,k,1)$ design $\mathcal{D}$ with point set V and we want to construct a 2-$(vk,k,1)$ design. (The parameters work.) We begin by taking k disjoint copies of $\mathcal{D}$, which means we have kb blocks and so we are short by

$$\frac{vk(vk-1)}{k(k-1)} - k\frac{v(v-1)}{k(k-1)} = v^2.$$

So we can finish our construction if we can find a collection of v^2 k-sets, consisting of one point from each copy of V. It is not hard to verify that this set must be an $OA(v,k)$, and that any $OA(v,k)$ will work.

6.3 Affine Planes

We have the following bound.

6.3.1 Lemma. *If an $OA(n,k)$ exists, then $k \leq n+1$.*

Proof. The *graph* of an orthogonal array $OA(n,k)$ has the n^2 rows of the array as its vertices, two rows are adjacent if they agree on some coordinate. The rows that take the same value in the same column form a clique in the graph, and thus each column determines a set of n vertex-disjoint cliques of size n. We call this a parallel class. Since two rows agree on at most one coordinate, we can color each edge of the graph by the index of the column where the corresponding rows agree. The subgraph formed by the edges of a given color form a parallel class, and different parallel classes are edge disjoint. Now the number of edges in the graph is at most

$$\binom{n^2}{2}$$

and the number of edges in a parallel class is

$$n\binom{n}{2}$$

and therefore the number of parallel classes is at most

$$\frac{n^2(n^2-1)}{n^2(n-1)} = n+1. \qquad \square$$

We can shorten this proof by omitting the graph, but the graph will be needed again. You may prove that a coclique in the graph of an $OA(n,k)$ has size at most n; if the chromatic number of the graph is n then the $OA(n,k)$ can be extended to an $OA(n,k+1)$ by adding a column.

6.3.2 Theorem. *An $OA(n,n+1)$ is an affine plane of order n.*

Proof. We take the n^2 rows as point set and the n-cliques of the graph as the lines. By our argument above, any two distinct points lie in exactly one line. Thus our points and lines form a 2-$(n^2,n,1)$ design and any such design is an affine plane. $\square$

In fact we see that an $OA(n,n+1)$ is, as an incidence structure, the dual of an affine plane. Any $OA(n,k)$ can be viewed as an incidence structure: the rows are the lines, the points are ordered pairs consisting of a column and a symbol and the point (i,a) is incident with the rows that have a in the i-th position. Our graph above is the block graph of the incidence structure.

The graph of an orthogonal array does not determine the array in general—all affine planes of order n give rise to the complete graph K_{n^2}. Of course if we keep track of the coloring, then we can reconstruct the array.

6.4 Block Graphs

We construct a graph from an orthogonal array $OA(n,k)$ by taking the rows as vertices and declaring that two rows are adjacent if they agree on some coordinate. The resulting graph on n^2 vertices is regular with valency $k(n-1)$. The n rows with i-th coordinate equal to α form a clique of size n, and so the i-th coordinate gives rise to n vertex-disjoint cliques of size n. Thus they form a subgraph isomorphic to nK_n and we obtain a set of k edge-disjoint copies of nK_n. We will call a copy of nK_n a *parallel class*.

Since the block graph of an $OA(k,n)$ contains cliques of size n, its chromatic number is at least n. The problem of deciding when it is exactly

n is very interesting. As the block graph contains n pairwise vertex-disjoint n-cliques, we see that a *coclique* (i.e., independent or stable set) must have size at most n, since it cannot contain two vertices from the same clique. Therefore our block graph is n-colourable if and only if it has n pairwise vertex-disjoint cocliques of size n.

If an $OA(k, n)$ consists of k columns from an $OA(k+1, n)$, we say it is *extendible*.

6.4.1 Theorem. *The block graph of an $OA(k, n)$ is n-colourable if and only the array is extendible.*

Proof. The block graph of an $OA(k+1, n)$ is the edge-disjoint union of $k+1$ copies of nK_n. If we delete one of these parallel classes we are left with the block graph of an $OA(k, n)$, and the deleted parallel class gives rise to n pairwise vertex-disjoint cocliques. Hence the block graph of the $OA(k, n)$ is n-colourable. The converse is left as an exercise. □

6.4.2 Lemma. *The block graph of a Latin square of order n is n-colourable if and only if it contains a coclique of size n.* □

6.4.3 Lemma. *If the Latin square L is the multiplication table of a cyclic group of even order, its block graph does not contain a coclique of size n.* □

6.5 Existence

The fundamental result is due to Chowla, Erdős and Straus:

6.5.1 Theorem. *Suppose k is a positive integer. There is an integer N_k such that if $n \geq N_k$, then there is an orthogonal array $O(k, n)$.* □

For a proof of this, see Chapter 22 in Van Lint and Wilson "A Course in Combinatorics". There is also a treatment in Section X.5 of Volume 2 of "Design Theory" by Beth, Jungnickel and Lenz. The basic strategy of the proof is to consider the set S_k of positive integers n such that an $OA(k, n)$ exists. This set contains all sufficiently large prime powers. Further each construction of arrays from smaller arrays leads to a closure operation on S_k, and the essence of the proof is to show that, given the known elements

of S_k and the closure operations that apply, it follows that S_k contains all sufficiently large integers.

In Section 5.5 of Ian Anderson's "Combinatorial Designs" you will find a proof that there is no $OA(4, 6)$, from which it follows that there is no affine (or projective) plane of order six.

Chapter 7

1-Factorizations

A *1-factor* of a graph X is a perfect matching. A *1-factorization* is a partition of $E(X)$ into edge-disjoint 1-factors. If X has a 1-factorization, then X must be regular on an even number of vertices. The Petersen graph shows that the converse fails. Each regular bipartite graph has a 1-factorization (since a regular bipartite graph with positive valency must have a 1-factor). You might show that a 1-factorization of the complete bipartite graph is equivalent to an $n \times n$ Latin square. In this chapter we focus on 1-factorizations of complete graphs.

7.1 Factorizing K_n

We construct a 1-factorization for K_n when n is even. Assume $n = 2m$ and take the vertices to be ∞ together with the elements of $\mathbb{Z}_{2n-1}$. Colour ∞i with i (for each i), if x and y are finite, colour xy with $n(x+y)$. This works because if $n(x + y) = n(x + z)$, then

$$x + y = 2n(x + y) = 2n(x + z) = x + z$$

and therefore $y = z$.

We can present this 1-factorization as array L with rows and columns indexed by $\mathbb{Z}_{2m-1}$, such that $L_{i,j}$ is the colour of the edge ij. Clearly this array is symmetric and the diagonal is undefined, but it is natural to set $L_{i,i} = i$. You might now check that this array is a Latin square.

A Latin square L of order n is *idempotent* if $L_{i,i} = i$ for each i. The Latin square we just constructed is idempotent.

7.1.1 Lemma. *The 1-factorizations of K_{2m} are in bijection with idempotent symmetric Latin squares of order $2m-1$.* □

We can get 1-factorizations of K_{4n} by combining two 1-factorizations of K_{2m} with a 1-factorization of $K_{2n,2n}$.

Exercise: The block graph of a Latin square of order n contains a coclique of size n if and only if we can permute its columns so that the resulting square is idempotent.

Exercise: if n is even there is no symmetric idempotent $n \times n$ Latin square.

7.2 Squares and Triples

Given a Steiner triple system with point set V of size v, we can define a $v \times v$ Latin square as follows. If $i, j \in V$ and $i \neq j$, define the product $i \star j$ to be the third point in the unique triple that contains i and j. Define $i \star i$ to be i. THen $\star$ is a commutative product on V, and its multiplication table is a symmetric idempotent Latin square.

Since symmetric idempotent $v \times v$ Latin squares exist for all odd v, and since Steiner triple systems can exist only when $v \equiv 1, 3$ modulo 6, it is clear that not every symmetric idempotent Latin square gives rise to a Steiner triple system. But given a symmetric idempotent $n \times n$ Latin square, we can construct a Steiner triple system on $v = 3n$ points as follows.

Let $N = \{1, \ldots, n\}$ and let V be the Cartesian product $V \times \mathbb{Z}_3$. Assume L is our Latin square as described. We start by taking all triples of the form

$$\{(r, i), (s, i), (L_{r,s}, i+1)\}$$

where r, s range over the elements of N and $r \neq s$, and $i \in \mathbb{Z}_3$ (and so the addition in the third coordinate is carried out mod 3). This gives us $3n(n-1)/2$ triples; adjoin to this set of triples all n triples

$$\{(r, 0), (r, 1), (r, 2)\}, \quad r \in N.$$

It is easy to verify that the resulting set of triples is a Steiner triple system, and so we have proved that if $v \equiv 3$ modulo 6, there is a Steiner triple system on v points.

Latin squares can also be used to provide an efficient construction of Steiner triple systems on v points when $v \equiv 1$ modulo 6. Suppose L is

a symmetric idempotent Latin square of order n, with entries from N as above. Let M be the Latin square we obtain from L by adding n to each entry. Then the array

$$\begin{pmatrix} L & M \\ M & L \end{pmatrix}$$

is a symmetric $2n \times 2n$ Latin square with diagonal entries

$$1, \ldots, n, 1, \ldots, n.$$

This square can be used to construct a Steiner triple system with points set consisting of a new symbol ∞ together with three disjoint copies of $\{1, \ldots, 2n\}$. We leave the construction as an exercise.

Chapter 8

Distance-Regular Graphs

Many intersting combinatorial and geometric striuctures are associated with interestin graphs where, in this context, interesting means "distance regular". We explain what this means.

8.1 Distance-Regular Graphs

A graph X is *distance regular* if, given a pair of vertices (a, b) at distance k, the number of vertices x at distance i from a and j from b is determined by i, j and k. We denote this number by $p_{i,j}(k)$, this set of constants forms the *intersection parameters* of the graph X. Primitive strongly regular graphs are distance regular. Cycles are distance regular.

Incidence graphs of symmetric designs are distance-regular with diameter three. Conversely a bipartite distance-regular graph with diameter three is the incidence graph of a symmetric design. Incidence graphs of generalized quadrangles are distance-regular if they are regular.

If $\mathcal{P}$ is a projective plane of order n, its incidence graph is distance-regular with diameter three and girth six. It is regular with valency $q + 1$ and has $2(q^2 + q + 1)$ points. Choose a line ℓ and point p on ℓ. If we delete from the incidence graph the vertices corresponding to the $q + 1$ points on ℓ and the vertices corresponding to the $q + 1$ lines on p. the resulting graph is bipartite (still) with diameter four. You may show that it is distance-regular.

We can also construct distance-regular graph from Hadamard matrices. Suppose H is Hadamard of order n. We construct a bipartite graph on $4n$

vertices by specifying its adjacency matrix. First we convert H to a matrix $\widehat{H}$ of order $2n \times 2n$ by replacing each entry 1 and -1 respectively by

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

Let X be the bipartite graph with adjacency matrix

$$\begin{pmatrix} 0 & \widehat{H} \\ \widehat{H}^T & 0 \end{pmatrix}.$$

Then X is distance regular with valency n.

A graph X of diameter d is *antipodal* if the relation "is equal to or is at distance d from" is an equivalence relation on $V(X)$. The cube is antipodal of diameter three. The line graph of the Petersen graph is antipodal of diamter three, and its antipodal classes have size three. The graphs we constructed from projective planes above, and the graphs we constructed from Hadamard matrices are antipodal. (In the projective case the antipodal classes have size n, in the Hadamard case they have size 2.)

Suppose X is a connected graph with diameter d. We define the i-th *distance graph* of X to be the graph with vertex $V(X)$, where two vertices are adjacent in X_i if and only if they are at distance i in X. We use A_i to denote the adjacency matrix of X_i and set $A_0 = I$. Of course $X_1 = X$. We note that $\sum_i A_i = J$.

The matrices $\{A_0, \ldots, A_d\}$ span a real vector space of dimensioon $d+1$. We say that X is a *distance-regular graph* if this vector space is closed under matrix multiplication, thus it is a matrix algebra, known as the *Bose-Mesner algebra* of the graph. (We note that since the set $\{0, A_0, \ldots, A_d\}$ is closed under Schur multiplication, our vector space is closed under Schur multiplication too.) It follows that if X is distance regular, there are constants $p_{i,j}(k)$ such that

$$A_i A_j = \sum_{r=0}^{R} p_{i,j}(r) A_r.$$

The parameters $p_{i,j}(k)$ are non-negative integers, and are known as the *intersection numbers* of the graph. Since the product of two symmetric matrices A and B is symmetric if and only if $AB = BA$, we see that if X is distnace-regular, then its Bose-Mesner algebra is commutative.

The intersection numbers have a straightforward combinatorial definition. If u and v are vertices of X at distance r (in X), then $p_{i,j}(r)$ is equal

to the number of vertices w of X such that $\text{dist}(w,u) = i$ and $\text{dist}(w, v = j)$. (We could define a distance-regular graph to be a graph where the intersection numbers are well-defined; this is in fact the usual definition.)

Suppose X is connected graph of diameter d. We say that X is *distance transitive* if for each r such that $0 \leq r \leq d$, the automorphism group of X acts transitively on the ordered pairs of vertices at distance r. Many of the distance-regular graphs we will meet are distance-transitive. Verifying that a graph is distance-transitive is often the easiest way to show that it is distance regular.

A distance-regular graph X is *primitive* if each of its distance graphs is connected. We note that a graph X is bipartite if and only if X_2 is not connected, whence bipartite distnce-regular graphs are imprimitive. For the d-cube we see that X_d is not connected; a distance-regular graph of diameter d is said to be *antipodal* if X_d is not connected. The line graph of the Petersen graph is distance-regular and antipodal but not bipartite. It is an important result that an imprimitive distance-regular graphs is bipartite or antipodal (or both).

If X is distance-regular with diameter d and antipodal, then X_d is a disjoint union of complete graphs.

8.2 An Example: Projective Planes

The incidence graph X of a projective plane is distance-regular. We verfiy this. Assume that the order of the plane is n and let N be its incidence matrix. Then

$$A_1 = \begin{pmatrix} 0 & N \\ N^T & 0 \end{pmatrix}$$

and so

$$A_1^2 = \begin{pmatrix} NN^T & 0 \\ 0 & N^TN \end{pmatrix} = nI + \begin{pmatrix} J & 0 \\ 0 & J \end{pmatrix}$$

Thus X_2 consists of two copies of the complete graph on $n^2 + n + 1$ vertices. You should now verify that

$$A_3 = \begin{pmatrix} 0 & J - N \\ J - N^T & 0 \end{pmatrix},$$

from which it follows that X is distance regular with diameter three.

You should also prove that if X is bipartite and distance regular with diameter three, then it is the incidence graph of a symmetric design. (The design is a projective plane if and only if the girth of the incidence graph is six.)

The incidence graph of a classical projective plane formed by the 1- and 2-dimensional subspaces of a vector space over a finite field is necessarily distance transitive. However for non-classical projective planes, the incidence graph is not distance transitive in general.

8.3 Partial Geometries

A *partial geometry* is a point and line-regular partial linear space with the property that there is a positive integer α such that each point not on a line is collinear with exactly α points on the line. We write $PG(s, t, \alpha)$ to denote a partial geometry with lines of size $s + 1$ and $t + 1$ lines on each point, where a point not on a line is collinear with exactly α points on the line. So an $OA(n, k)$ is a $PG(k - 1, n - 1, k - 1)$. A 2-design with $\lambda = 1$ is a $PG(k - 1, r - 1, k)$. (I am sorry about all the -1's, but the geometers got here first.)

You can prove the following:

(a) A $PG(s, t, s + 1)$ is equivalent to a 2-design with parameters $(st + s + 1, s + 1, 1)$.

(b) A $PG(s, t, s)$ is equivalent to an orthogonal array $OA(t + 1, s + 1)$.

(c) The dual of a partial geometry $PG(s, t, \alpha)$ is a $PG(t, s, \alpha)$.

(d) The edges and 1-factors of K_6 form a $PG(2, 2, 1)$.

(e) An incidence structure is a $PG(s, t, 1)$ if it is point and block regular and its incidence graph has diameter four and girth eight.

There are partial geometries where

$$1 < \alpha < \min\{s, t\}$$

but they are not easy to find. (They are sometimes said to be *proper*.)

8.3.1 Lemma. *Let N be a 01-matrix. Then N is the incidence matrix of a partial geometry if and only if there are positive integers s, t and α such that*

(a) $N\mathbf{1} = (t+1)\mathbf{1}$.

(b) $N^\top \mathbf{1} = (s+1)\mathbf{1}$.

(c) $NN^TN = (t+s+1)N + \alpha(J-N) = (s+t+1-\alpha)N + \alpha J$. □

We leave the proof of this as an exercise. If our partial geometry has v points and b lines, then we have

$$(t+1)^2(s+1)\mathbf{1} = (t+s+1-\alpha)(t+1) + \alpha b$$

whence we obtain

$$b = (t+1)\frac{st+\alpha}{\alpha}, \quad (s+1) = \frac{st+\alpha}{\alpha}.$$

(The expression for v is a consequence of the expression for b, in more than one way.)

A partial geometry is a partial linear space and therefore the matrix $NN^T - (t+1)I$ is a 01-matrix. Hence it is the adjacency matrix of the point graph of the geometry and $N^TN - (s+1)I$ is the adjacency matrix of its line graph. If A and B are matrices such that both products AB and BA are defined (i.e., A and B^T have the same order), then

$$\det(I + xAB) = \det(I + xBA)$$

From this it follows that NN^T and N^TN have the same non-zero eigenvalues with the same multiplicities. Now from Lemma 8.3.1(c) we get that

$$(NN^T)^2 = (s+t+1-\alpha)N + \alpha(s+1)J$$

and from this it follows that the eigenvalues of NN^T are

$$0, \quad s+t+1-\alpha, \quad (s+1)(t+1)$$

with respective multiplicities $v - \mathrm{rk}(N)$, $\mathrm{rk}(N) - 1$ and 1. Consequently the eigenvalues of the point graph are

$$-t-1, \ s-\alpha, \ s(t+1)$$

and those of the line graph are

$$-s-1,\ t-\alpha,\ (s+1)t.$$

In both these case the third eigenvalue is the valency of the graph, and is simple, while the multiplicity of the second eigenvalue is equal to $\mathrm{rk}(N)-1$.

8.4 Strongly Regular Graphs

A graph X is *strongly regular* if it is neither complete nor empty and there are integers k, a and c such that

(a) X is regular with valency k.

(b) Any two adjacent points have exactly λ common neighbors.

(c) Any two points that distinct and not adjacent have exactly c common neighbours.

If X is a strongly regular graph on v vertices its parameter vector is $(v, k; a, c)$.

If $m, n > 1$ then the disjoint union mK_n of m copies of K_n is strongly regular with $c = 0$; its complement is strongly regular. The complement of a strongly regular is strongly regular. A strongly regular graph is *imprimitive* if it is not connected (in which case it is mK_n), or if its complement is not connected (in which case it is $\overline{mK_n}$).

8.4.1 Lemma. *Let A be the adjacency matrix of the graph X. The X is strongly regular with parameters $(v, k; a, c)$ if and only if*

$$A^2 - (a-c)A - (k-c)I = cJ.$$

The block graph of a 2-$(v, k, 1)$ design is strongly regular or complete. The simplest way to prove this is to show to compute k, a and c. However this fact is a corollary of the following:

8.4.2 Lemma. *The points and block graphs of a partial geometry are strongly regular.* □

This lemma is an easy consequence of the eigenvalue computations above, Lemma 8.4.1 and the following:

8.4.3 Lemma. *A connected graph is strongly regular if and only if its adjacency matrix has exactly three distinct eigenvalues.* □

A connected regular graph whose adjacency matrix has exactly two eigenvalues must be a complete graph. (The number of distinct eigenvalues, less one, is an upper bound on the diameter of a graph.)

Constructions: symmetric Hadamard matrices with constant row sums and constant diagonal. Block graphs of orthogonal arrays and 2-$(v, k, 1)$ designs. Symmetric conference matrices with constant diagonal. Complements of srgs.

8.5 Generalized Quadrangles

An incidence structure is a *generalized quadrangle* if:

(a) Any two points lie on at most one line.

(b) If p is a point of the line ℓ, there is a unique point on ℓ that is collinear with p.

We generally assume that a GQ (generalized quadrangle) is not a line or a dual line. (These cases are *degenerate*.) The dual of a GQ is a GQ, as you should verify.

We offer a somewhat surprising example. Let $\mathcal{P}$ be the set of 15 unordered pairs from $\{0, 1, 2, 3, 4, 5\}$, viewed as the edges of the complete graph K_6. Let $\mathcal{F}$ denote the 15 1-factors in K_6. Then the incidence structure $(\mathcal{P}, \mathcal{F})$ (with inclusion as incidence) is a generalized quadrangle.

For an actual family, let $(\mathcal{P}, \mathcal{L})$ be the incidence structure with the vertices of the complete bipartite graph $K_{m,n}$ as points and its edges as lines. These are again GQs, but form a trivial class known as *dual grids*. (As you might hope, the dual of dual grid is a *grid*.)

As noted earlier the points and lines of polar space of rank two form a generalized quadrangle, and these provide some of the most important classes. We will meet others.

8.5.1 Lemma. *A thick finite generalized quadrangle is point regular and line regular.* □

The convention is that by $GQ(s,t)$ we denote a generalized quadrangle with $s+1$ points on each line and $t+1$ lines on each point. Our first example above is a $GQ(2,2)$. A grid is a $GQ(s,1)$ (and so a dual grid is a $GQ(1,t)$).

8.5.2 Lemma. *Suppose $\mathcal{P}$ is a partial linear space that contains non-collinear points and nonconcurrent lines. Then $\mathcal{P}$ is a generalized quadrangle if and only if its incidence graph has diameter four and girth eight.* □

8.5.3 Theorem. *The point graph of a $GQ(s,t)$ is strongly regular with parameters*

$$((s+1)(st+1), s(t+1), s-1, t+1).$$

Proof. Suppose x and y are two points that are not collinear. Then y lies on $t+1$ lines and on each of these lines there is a unique point collinear with x, hence x and y have exactly $t+1$ common neighbours. If x and y are distinct collinear points, there are $s-1$ points on $x \vee y$ collinear with both x and y and no point off $x \vee y$ is collinear with x and y. Therefore two adjacent points have exactly $s-1$ common neighbours.

Since any point is on $t+1$ lines, each containing a further s points, the valency of our graph is $s(t+1)$.

Finally we calculate the number of vertices. Choose a line ℓ. Each of the points on ℓ lies on a set of t lines distinct from ℓ, and these sets of t lines are pairwise disjoint. Since each point off ℓ is collinear with a unique point on ℓ, we find that the number of vertices is

$$s+1+(s+1)ts = (s+1)(st+1). \qquad \square$$

The eigenvalues of the point graph are the valency $s(t+1)$ and the two zeros of

$$x^2-(s-t-2)x-(s-1)(t+1) = (x+t+1)(x-s+1),$$

i.e., they are $s-1$ and $-t-1$.

8.5.4 Lemma. *Let N be the incidence matrix of a $GQ(s,t)$ and let A be the adjacency matrix of its point graph. Then*

$$NN^TN = (s+t)N + J.$$

Proof. We have

$$NN^T = (t+1)I + A$$

where A is the adjacency matrix of the point graph. Then

$$NN^TN = (t+1)N + AN.$$

If x is a point and ℓ is a line, then $(AN)_{x,\ell}$ is the number of points w collinear with x on ℓ. There are two cases. If $x \in \ell$ then this number is s. If $x \notin \ell$, it is 1. Hence

$$AN = sN + (J - N)$$

and the lemma follows. □

Since $AN = (s-1)N + J$, if $\mathbf{1}^T z = 0$ then

$$ANz = (s-1)Nz$$

and so the differences of distinct columns of N are eigenvectors for A with eigenvalue $s-1$. Since $A+(t+1)I = NN^T$, we see that the non-zero vectors in $\ker(N^T)$ are eigenvectors for A with eigenvalue $-t-1$. It follows that the multiplicity of $s-1$ as an eigenvalue if $\mathrm{rk}(N)-1$ and the multiplicity of $-t-1$ is $v - \mathrm{rk}(N)$. The multiplicities of $s-1$ and $-t-1$ are respectively

$$\frac{st(s+1)(t+1)}{s+t}, \quad \frac{s^2(st+1)}{s+t}$$

This implies that $s+t$ must divide $st(st+1)$, which is a non-trivial constraint on the possible values of s and t.

[***This all follows from the computations with partial geometries, should be deleted***]

Grids and dual grids aside, all known (finite) GQs have parameters of the form

$$(q,q),\ (q,q^2),\ (q^2,q),\ (q^2,q^3),\ (q^3,q^2),\ (q-1,q+1),\ (q+1,q-1)$$

where q is a prime power.

8.6 Higman's Inequality

The following inequality is due to D. Higman, the proof we present is due to Cameron.

8.6.1 Theorem. *If a $GQ(s,t)$ exists then $t \leq s^2$.*

Proof. Let x and y be two noncollinear points in our GQ, let C denote $x^\perp \cap y^\perp$, and let N be the set of points that are not collinear with x or y. We count ordered pairs (w,u) in $N \times C$ where $u \sim w$. Then

$$\sum_{w \in N} |w^\perp \cap C| = (t+1)(t-1)s.$$

Next count ordered triples in $N \times C \times C$ to get

$$\sum_{w \in N} |w^\perp \cap C|(|w^\perp \cap C| - 1) = (t+1)t(t-1).$$

We have

$$m = s^2t - st - s + t$$

and if we set μ equal to the average value of $|w^\perp \cap C|$, then by Jensen's inequality

$$\frac{1}{m}(t+1)t(t-1) \geq \mu(\mu - 1).$$

After some manipulation this yields that

$$t(s-1)(s^2 - t) \geq 0. \qquad \square$$

If $t = s^2$ then $|w^\perp \cap C| = s + 1$ (as you might show).

Suppose X is a strongly regular graph with parameters $(n, k; a, c)$. Let ℓ denote the valency $n - 1 - k$ of the complement of X and let θ and τ be the eigenvalues of X distinct from k. The *Krein condition* assures us that

$$1 + \frac{\lambda^3}{k^2} - \frac{(\lambda+1)^3}{\ell^2} \geq 0.$$

If we apply this to the point graph of a $GQ(s,t)$ with $\lambda = -t - 1$, we get that

$$0 \leq 1 - \frac{(t+1)^3}{s^2(t+1)^2} + \frac{t^3}{s^4t^2} = \frac{s^4 - s^2 - (s^2-1)t^2}{s^4}$$

which yields Higman's inequality.

8.7 Eigenvalues of Neighborhoods in SRGs

Let X be the point graph of a $GQ(s,t)$. This is a strongly regular graph. The neighbourhood of a vertex consists of $t+1$ complete graphs of size s, so its eigenvalues are $s-1$ and -1 with respective multiplicities $t+1$ and $(t+1)(s-1)$.

Somewhat surprisingly it is possible to determine the eigenvalues of the second neighborhood of a vertex of a strongly regular graph—the subgraph induced by the vertices at distance two from the vertex—in terms of the eigenvalues of the neighborhood.

Proof. We assume the parameters of X are $(n,k;a,c)$. We can write the adjacency matrix A of X in partitioned form:

$$A = \begin{pmatrix} 0 & \mathbf{1}^T & 0 \\ \mathbf{1} & A_1 & B^T \\ 0 & B & A_2 \end{pmatrix}$$

Our aim is to determine the eigenvalues of A_2.

The 3-dimensional space spanned by vectors constant on the distance partition relative to the vertex 1 is A-invariant; the eigenvalues of A on this subspace are k, θ and τ. The orthogonal complement of this subspace is also A-invariant, and consists of vectors that are zero on vertex 1 and sum to zero on the first and second neighborhoods. Let us denote this space by V_0. Our goal is to find eigenvectors for A in V_0.

We note that the minimal polynomial of A on V_0 is $(t-\theta)(t-\tau)$, in particular it is a quadratic polyomial and therefore if $w \in V_0$ then the space spanned by w and Aw is A-invariant. Thus either it is 1-dimensional and w is an eigenvector for A (with eigenvalue θ or τ), or it is 2-dimensional and is spanned by eigenvectors for A, one with eigenvalue θ and with eigenvalue τ.

We start by choosing an eigenvector z for A_1 that is orthogonal to $\mathbf{1}$. Assume that $A_1 z = \lambda z$ and set

$$\hat{z} = \begin{pmatrix} 0 \\ z \\ 0 \end{pmatrix}, \qquad \hat{w} = \begin{pmatrix} 0 \\ 0 \\ Bz \end{pmatrix}.$$

Then $\hat{z} \in V_)$ and

$$A\hat{z} = \begin{pmatrix} 0 \\ A_1 z \\ Bz \end{pmatrix} = \lambda \hat{z} + \hat{w}$$

and

$$A\hat{w} = \begin{pmatrix} 0 \\ B^T Bz \\ A_2 Bz \end{pmatrix}.$$

Since Aw must be a linear combination of $\hat{z}$ and $\hat{w}$, there must be scalars β and μ such that

$$B^T Bz = \beta z, \qquad A_2 Bz = \mu Bz$$

and therefore

$$Aw = \beta\hat{z} + \mu\hat{w}.$$

So, relative to the basis $\{\hat{z}, \hat{w}\}$ the effect of A is represented by

$$\begin{pmatrix} \lambda & \beta \\ 1 & \mu \end{pmatrix}.$$

The eigenvalues of this matrix are θ and τ, whence

$$\lambda + \mu = \theta + \tau = a - c, \quad \lambda\mu - \beta = \theta\tau = c - k.$$

The first conclusion we reach is that if z is an eigenvector for A_1 with eigenvalue λ and $Bz \neq 0$, then Bz is an eigenvector for A_2 with eigenvalue $a - c - \lambda$.

If X is a regular graph and θ is an eigenvalue with eigenvector orthogonal to $\mathbf{1}$, then $-\theta - 1$ is an eigenvalue of X. Also if X is connected and k-regular, then k is a simple eigenvalue with $\mathbf{1}$ as an eigenvector.

Hence if there is an eigenvector w for A_2 with eigenvalue μ, then w is an eigenvector for $\overline{A^T}$ and if $(J - B^T)w \neq 0$, by the argument above it is an eigenvector for $\overline{A_1}$ with eigenvalue

$$-\theta - \tau - 2 - (-\mu - 1) = -\theta - \tau + \mu - 1$$

and therefore it is an eigenvector for A_1 with eigenvalue $\theta + \tau - \mu$. (Note that all row sums of B are equal, as are all column sums, so $(J - B^T)w = 0$ if and only if $B^T w = 0$.)

Since BB^T is positive semidefinite, $\beta \geq 0$. As λ and μ are roots of

$$(t - \lambda)(t - \mu) = t^2 - (\lambda + \mu) + \lambda\mu = t^2 - (\theta + \tau)t + \theta\tau + \beta$$

we see that

$$\tau \leq \lambda, \mu \leq \theta.$$

If $A_1 z = \theta z$ then $\lambda = \theta$ and $\mu = \tau$ and

$$\beta = \lambda\mu - \theta\tau = 0.$$

Therefore $Bz = 0$. Similarly $Bz = 0$ if $\lambda = \tau$.

8.8 Eigenvalues of Neighborhoods in GQs

We apply the results of the previous to section to the point graph of a $GQ(s,t)$. The neighborhood of a vertex in such a graph consists of $t+1$ vertex disjoint cliques of size s. Hence the eigenvalues of the neighborhood are $s-1$ (with multiplicity $t+1$) and -1 (with multiplicity $(t+1)(s-1)$). Since $s-1$ is an eigenvalue of the GQ, each eigenvector for the neighborhood with eigenvalue $s-1$ that sums to zero gives rise to an eigenvector of the GQ that is supported on the neighborhood.

Now consider the second neighborhood. It is connected with valency

$$k - c = s(t+1) - (t+1) = (s-1)(t+1).$$

The other possible eigenvalues are $s-1$ and $-t-1$ (eigenvalues of the GQ, with respective multiplicities f and g say) and

$$a - c - (-1) = s - t - 2 + 1 = s - t - 1$$

with multiplicity $(s-1)(t+1)$. Then

$$s^2 t = 1 + (s-1)(t+1) + f + g$$

and, since $\mathrm{tr}(A_2) = 0$,

$$0 = (s-1)(t+1) + (s-1)(t+1)(s-t-1) + f(s-1) - g(t+1).$$

These equations yield

$$f + g = s^2 t - st - s + t$$
$$f(s-1) - g(t+1) = (s-t)(s-1)(t+1)$$

and after some effort we deduce that

$$f = \frac{s^2(t^2-1)}{s+t}, \quad g = \frac{t(s^2-t)(s-1)}{s+t}.$$

We summarize our conclusions:

multiplicity:	1	$\frac{s^2(t^2-1)}{s+t}$	$(s-1)(t+1)$	$\frac{t(s^2-t)(s-1)}{s+t}$
eigenvalue:	$(s-1)(t+1)$	$s+1$	$s-t-1$	$-t-1$

This provides another proof of the inequality $t \leq s^2$; we also see that if equality holds then the second neighborhood has exactly three eigenvalues and therefore it must be strongly regular.

Chapter 9

Block Intersections

9.1 Quasi-Symmetric Designs

A design is *quasi-symmetric* if there are distinct integers x and y such that any two distinct blocks have either x or y points in common. Any 2-$(v, k, 1)$ proivides an example where $x = 0$ and $y = 1$. We usually assume that $x < y$. We define the *block graph* to be the graph with the blocks as its vertices, where two blocks are adjacent if and only if they intersect in exactly y points.

The first class of examples is the least interesting: if $m > 1$, take m copies of each of the blocks of a symmetric design. We call this a *multiple of a symmetric design*, it is quasi-symmetric with $x = \lambda$ and $y = k$. It easy to see that any quasi-symmetric design with $y = k$ must be a multiple of a symmetric design.

The second class consists of the 2-$(v, k, 1)$ designs, here we have $x = 0$ and $y = 1$. Conversely, any quasi-symmetric design with $x = 0$ and $y = 1$ is a 2-$(v, k, 1)$ design.

Thirdly we have the so-called *strongly resolvable designs*. A design is strongly resolvable if there is a partition of its blocks into classes and constants ρ and μ such that any two distinct blocks in the same class intersect in ρ points, while two blocks in disjoint classes meet in μ points. These are quasi-symmetric with $x = \rho$ and $y = \mu$, and can be characterized as the quasi-symmetric designs with $x = k + \lambda - r$. (The claims in this last sentence are not obvious, but will proved in **??** and **??**.) The simplest examples of strongly resolvable designs are the affine planes. The block graph

of a strongly resolvable design is complete multipartite graph.

The fourth class may be finite. Let $\mathcal{D}$ be a symmetric design and let β be a fixed block in it. The *residual design* has the points of $\mathcal{D}$ not in β as its point set, and the intersection of this set with the blocks of $\mathcal{D}$ (other than β) as its blocks. Any residual design has $r = k + \lambda$; designs for which this condition holds are sometimes known as *quasi-residual*. The residual design of a symmetric design where $\lambda = 2$ can be shown to be quasi-symmetric with degree set $\{1, 2\}$. A symmetric design such that $\lambda = 2$ is 'dignified' by the name *biplane*. It is an open question as to whether there are infinitely many biplanes. (Or, more generally, whether there are infinitely many symmetric 2-(v, k, λ) designs for any value of λ greater than one.) Any 2-design with the parameters of a residual biplane, that is with $\lambda = 2$ and $r = k+2$, must be the residual design of a biplane (see ???).

The fifth class is finite, with cardinality four. It consists of the Witt designs on 22 and 23 points, and their complements.

Suppose B is the incidence matrix of a quasi-symmetric 2-(v, k, λ) with degree set $\{x, y\}$. Then

$$BB^T = (r - \lambda)I + \lambda J,$$
$$B^T B = (k - x)I + (y - x)A + xJ.$$

As BB^T and $B^T B$ have the same non-zero eigenvalues with the same multiplicities, we can use these identities to determine the spectrum of A. Because B has constant row and column sum, $B^T B$ must commute with J, hence A must commute with J.

We have

$$kr\mathbf{1} = B^T B\mathbf{1} = (k - x)\mathbf{1} + (y - x)A\mathbf{1} + xv\mathbf{1},$$

whence

$$A\mathbf{1} = (y - x)^{-1}(kr - k + x - xv)\mathbf{1}.$$

This gives one eigenvalue and eigenvector for A.

9.2 Triangle-free Strongly Regular Graphs

A graph is strongly regular with parameters $(n, k; a, c)$ if it has n vertices, valency k, any two adjacent vertices have exactly a common neighbours and

any two distinct non-adjacent vertices have exactly c common neighbours. The complement of a strongly regular graph is strongly regular; in fact we could define strongly regular graphs to be the graphs arising as colour classes in association schemes with two classes. The disjoint union mK_n of m copies of K_n is strongly regular provided $m > 1$ and $n > 1$—the complete and empty graphs are not usually considered to be strongly regular. A strongly regular graph G is *primitive* if both G and its complement are connected; the only imprimitive strongly regular graphs are the graphs mK_n and their complements, the complete multipartite graphs.

The smallest non-trivial strongly regular graph is the pentagon C_5. The line graphs of the complete graphs K_n and the complete bipartite graphs $K_{n,n}$ are as well. Hence the Petersen graph is strongly regular. A strongly regular graph is *triangle-free* if it has no triangles, which is the same as requiring that $a = 0$. Only seven primitive triangle-free strongly regular graphs are known (and you have just met two of them). One can be constructed from the Witt design on 22 points as follows.

This design has parameters 3-$(22, 6, 1)$; hence it has 77 blocks. We construct a graph HS with vertex set consisting of all points and blocks of this design, and one extra point which we denote by ∞. The adjacencies are as follows. The vertex ∞ is adjacent to each of the 22 vertices corresponding to the points of the design. Each of these 22 vertices is in turn adjacent to the vertices representing the 21 blocks which lie on it. Each 'block vertex' is adjacent to the vertices representing the blocks disjoint from it. Although it is not obvious, this construction produces a vertex-transitive graph which is strongly regular with parameters $(100, 22; 0, 6)$. It is known as the *Higman-Sims graph*. (It contains, as induced subgraphs, strongly regular graphs on 16, 50 and 56 vertices—the Clebsch, Hoffman-Singleton and Gewirtz graphs respectively.)

This construction can be reversed in part. Suppose X is an $(n, k; 0, c)$ strongly regular graph, and let V denote the set of vertices adjacent to some fixed vertex u in X. We define an incidence structure with point set V and blocks consisting of the subsets of V with size c which have a common neighbour at distance two from u. It is not hard to show that this forms a 2-$(k, c, c-1)$ design with $k(k-1)/c$ blocks, possibly repeated, and $r = k-1$. Two adjacent vertices at distance two from u have no common neighbour adjacent to u; hence they determine disjoint blocks.

9.2.1 Lemma. *Let X be an $(n, k; 0, c)$ strongly regular graph, and let $\mathcal{D}$*

denote the design on the neighbours of some fixed vertex, formed as just described above. Then the following conditions are equivalent:

(a) $\mathcal{D}$ *is a 3-design,*

(b) $\mathcal{D}$ *is quasi-symmetric with* $x = 0$*,*

(c) $k = \frac{1}{2}[(3c+1) + (c-1)\sqrt{4c+1}]$*,*

(d) The graph induced by the vertices at distance two from a fixed vertex is strongly regular.

Proof. The design on the neighbours of the vertex u of X has parameters 2-$(k, c, c-1)$. As it has $k(k-1)/c$ blocks, it follows from ?? that (a) and (b) are equivalent.

If v is at distance two from u then it has exactly $k - c$ neighbours at distance two from u; therefore there are at least $k - c$ blocks disjoint from the block corresponding to v. By ??, we then have

$$\frac{k(k-1)}{c} \geq 1 + \frac{c(k-2)^2}{c^2 - 3c + k} + k - c$$

After some calculation (in Maple, preferably) we find that the difference between the two sides of this inequality is

$$(k-c)\frac{k^2 - 3kc - k + c + 4c^2 - c^3}{c(c^2 - 3c + k)};$$

from this we deduce that

$$k \geq \frac{1}{2}[3c + 1 + (c-1)\sqrt{4c+1}],$$

with equality if and only if $\mathcal{D}$ is quasi-symmetric (with $x = 0$). Hence (b) and (c) are equivalent.

The argument we just used shows that if equality holds in the previous equation, then there are exactly $k - c$ blocks disjoint from a given block in $\mathcal{D}$. It follows that two vertices at distance two from v are adjacent if and only if the corresponding blocks are disjoint, and therefore the graph $X_2(v)$ is the complement of the block graph of $\mathcal{D}$. As $\mathcal{D}$ is quasi-symmetric, it must be strongly regular. Thus (c) implies (d).

The size of the intersection of two distinct blocks of $\mathcal{D}$ is determined by the number of common neighbours of the corresponding vertices in $X_2(v)$. Therefore $\mathcal{D}$ is quasi-symmetric if and only if $X_2(v)$ is strongly regular. □

9.3 Resolvable Designs

A *parallel class* in a design is a collection of blocks that partitions the point set. A design is *resolvable* if its block set can be partitioned into parallel classes. The canonical example is the partition of the lines of an affine plane into parallel classes. A resolvable design where any two blocks in distinct classes meet in the same number of points is known as an *affine resolvable design*. Examples are provided by the points and lines of an affine geometry, and by the Hadamard 3-designs.

We have the following strengthening of Fisher's inequality.

9.3.1 Lemma. *Let $\mathcal{D}$ be a 2-design with b blocks. If there is a partition π of the blocks of $\mathcal{D}$ into 1-designs, then $b \geq v + |\pi| - 1$.*

Proof. Let B be the incidence matrix of $\mathcal{D}$ and let $R_1, \ldots, R_c$ be the classes of a partition of the blocks of $\mathcal{D}$ into 1-designs. As R_i is a 1-design, the sum of the columns of B corresponding to the blocks in R_i is a positive multiple of $\mathbf{1}$. Let B' be the $v \times (b - c + 1)$ matrix we get from B by deleting one column from each class, and then adding a column with each entry equal to one. By what we have just proved, B and B' have the same column space and therefore they have the same rank. Because $\operatorname{rk} B = v$, it follows that $b - c + 1 \geq v$ as required. □

It is natural now to ask what happens if equality holds here; we will prove that this can happen if and only if the cells of the partition are dual 2-designs. For this we will need the following result from linear algebra.

9.3.2 Lemma. *Let B be a matrix with linearly independent rows. Then the matrix representing orthogonal projection onto the column space of B^T is $B^T(BB^T)^{-1}B$.* □

We will need some consequences of this fact, the proofs of which are left as exercises. If B is the incidence matrix of a 2-design then the projection P onto the column space of B^T is given by

$$P = \frac{1}{r - \lambda}\left(B^T B - \frac{\lambda k}{r} J\right).$$

Next, suppose that $R_1, \ldots, R_c$ is a partition of the blocks of $\mathcal{D}$ and let Q be the $b \times c$ matrix whose i-th column is $|R_i|^{-1/2}$ times the characteristic

vector of the class R_i. The column space of Q is the space of functions on the blocks of $\mathcal{D}$ that are constant on the classes of the resolution and orthogonal projection onto this space is represented by the matrix QQ^T. The projection onto the orthogonal complement of $\mathbf{1}$ in this space is given by

$$M := QQ^T - \frac{1}{b}J.$$

9.3.3 Lemma. *If the P are are just defined, then $PM = MP = 0$.*

Proof. As $MP = (PM)^2$, we need only show that $PM = 0$. For this we show that $JM = 0$ and $BM = 0$. The former is left entirely up to you. For the latter, observe that QQ^T is block diagonal and the i-th block is equal to $\frac{1}{m}J$ (where m is the size of the i-th cell of $|\pi|$). The sum of the columns of B in the i-th cell of π is equal to

$$\frac{mk}{v}\mathbf{1}$$

and therefore

$$BQQ^T = \frac{k}{v}J = \frac{r}{b}J.$$

SInce $BJ = rJ$, we conclude that $BM = 0$. □

9.3.4 Theorem. *Suppose $\mathcal{D}$ be a 2-design whose block set can be partitioned into c 1-designs. If $b = v + c - 1$, two distinct blocks in the same class meet in exactly $r - \lambda - k$ points while blocks in distinct classes meet in k^2/v points.*

Proof. Let B be the incidence matrix of $\mathcal{D}$. Suppose $R_1, \ldots, R_c$ is a partition of the blocks of $\mathcal{D}$ into 1-designs and $c = b + 1 - v$. Let P be the matrix representing orthogonal projection onto the column space of B^T, and let M be as above. By the lemma, $PM = 0$ whence it follows that $P + M$ is also a projection matrix. Because its rank is $v + c - 1$, we find that $P + M$ is the identity matrix when $b = v + c - 1$. Accordingly $P + M = I$.

If α and β lie in different classes of our partition then

$$(M)_{\alpha,\beta} = -1/b.$$

On the other hand, if $|\alpha \cap \beta| = y$ then

$$(P)_{\alpha,\beta} = \frac{1}{r - \lambda}\Big(y - \frac{\lambda k}{r}\Big).$$

Since $P + M = I$ we see that $(P)_{\alpha,\beta} + (M)_{\alpha,\beta} = 0$, whence

$$y = \frac{\lambda k}{r} + \frac{r-\lambda}{b} = \frac{\lambda v r + r(r-\lambda)}{br} = \frac{\lambda v + r - \lambda}{b}$$

Because $\lambda(v-1) = r(k-1)$, this proves that $y = k^2/v$.

Assume next that α and β are distinct blocks in the same class of the partition, and that $|\alpha \cap \beta| = x$. Suppose that there are exactly m blocks in the class containing α and β. Then

$$(M)_{\alpha,\beta} = \frac{1}{m} - \frac{1}{b}, \qquad (P)_{\alpha,\beta} = \frac{1}{r-\lambda}\left(x - \frac{\lambda k}{r}\right),$$

from which it follows that

$$m\left(\frac{k^2}{v} - x\right) = r - \lambda.$$

Now count the pairs (i, γ) where $\gamma \in \mathcal{D}$ and $i \in \gamma \cap \alpha$; this yields

$$k + (m-1)x + (b-m)\frac{k^2}{m} = kr$$

and from this we find that

$$m\left(\frac{k^2}{v} - x\right) = k - x.$$

This implies that $x = k + \lambda - r$, and also that each cell of π has size m. □

This theorem shows that a 2-design which admits a partition into $b-v+1$ 1-designs must be quasisymmetric. Further each 1-design is a dual 2-design (which might have repeated blocks).

In the next section we will see that any two distinct blocks in a 2-design meet in at least $k + \lambda - r$ points.

9.4 Designs with Maximal Width

The *distance* between blocks α and β is $|\alpha \setminus \beta|$. The *width* of a design is the maximum distance between two blocks in it. Majumdar has shown that two distinct blocks in 2-design must have at least $k + \lambda - r$ points in common which provides an upper bound of $r - \lambda$ on the width.

We prove Majumdar's result.

9.4.1 Theorem. *Any two blocks in a 2-design have at least $k+\lambda-r$ points in common. If equality holds than the relation "meet in k or $k+\lambda-r$ points" is an equivalence relation on the blocks of the design.*

Proof. Let $\mathcal{D}$ be a 2-design with incidence matrix B, and let P be the orthogonal projection onto the column space of B^T. We saw that in the last section that

$$P = \frac{1}{r-\lambda}\left(B^TB - \frac{\lambda k}{r}J\right).$$

Let α and β be distinct blocks of $\mathcal{D}$ and suppose $|\alpha \cap \beta| = x$. Consider the 2×2 submatrix M of $(r-\lambda)(I-P)$ formed by the intersections of the rows and columns corresponding to α and β; it is equal to

$$\begin{pmatrix} r-\lambda-k+\frac{\lambda k}{r} & -x+\frac{\lambda k}{r} \\ -x+\frac{\lambda k}{r} & r-\lambda-k+\frac{\lambda k}{r} \end{pmatrix}.$$

Because $I-P$ is a projection it is positive semi-definite, and thus M is positive semi-definite as well. Hence $\det M \geq 0$, which implies that

$$r-\lambda-k+\frac{\lambda k}{r} \geq -x+\frac{\lambda k}{r}.$$

Thus we have proved that any two distinct blocks of $\mathcal{D}$ meet in at least $k+\lambda-r$ points.

We now focus on the situation when equality holds in this bound. As $I-P$ is positive semi-definite, it is the Gram matrix of a set of vectors in $\mathbb{R}^m$ (where $m = \mathrm{rk}(I-P)$). If $x = k+\lambda-r$ then all entries of M are equal, from which it follows that the α- and β-rows of $I-P$ must be equal. This shows that γ is a block in $\mathcal{D}$ other than α and β then

$$|\gamma \cap \alpha| = |\gamma \cap \beta|.$$

Thus we have proved that "meeting in k or $k+\lambda-r$ points" is an equivalence relation on the blocks of $\mathcal{D}$. □

Let $\mathcal{D}$ be a 2-design with degree set $\{x,y,z\}$, where $x<y<z$ and $x = k+\lambda-r$. By an exercise???, the number of blocks which meet a given block α of $\mathcal{D}$ in exactly i points is independent of α. Hence the number of blocks that meet α in exactly x points does not depend on our choice of α. Thus there is a partition of the blocks of $\mathcal{D}$ with c cells, all of size b/c, such

that the size of the intersection of two distinct blocks is determined by the cells in which they lie. In particular, two distinct blocks lie in the same cell if and only if they intersect in x points.

The *class graph* of $\mathcal{D}$ is defined to be the graph with the equivalence classes of $\mathcal{D}$ as its vertices, and with two vertices adjacent if and only a block in one class meets a block in the other class in y points. The class graph of $\mathcal{D}$ must be regular, but more is true. The *degree* of a design is the number of different values taken by $|\alpha \cap \beta$, where α and β run over the pairs of distinct blocks.

9.4.2 Lemma. *Suppose $\mathcal{D}$ is a 2-design with degree three. If $k+\lambda-r$ lies in its degree set of $\mathcal{D}$, then the class graph of $\mathcal{D}$ is strongly regular.*

Proof. Let $\{x, y, z\}$ be the degree set of $\mathcal{D}$. Assume $x < y < z$ and $x = k+\lambda-r$; let c denote the number of classes of $\mathcal{D}$. If B is the incidence matrix of $\mathcal{D}$ then, from our discussion above, we may write

$$B^T B = (k-x)I + M \otimes J_{b/c}$$

where

$$M = xI + yA + z(J - I - A) = (x-z)I + (y-z)A + zJ.$$

We will show that M has exactly three distinct eigenvalues. It follows that the same is true for A; since the class graph is regular this implies it must be strongly regular. (See, e.g., [] for this.)

The key observation is that the non-zero eigenvalues of $B^T B$ are the non-zero eigenvalues of BB^T. As

$$BB^T = (r-\lambda)I + \lambda J,$$

the eigenvalues of BB^T are kr (with multiplicity one) and $r-\lambda$ (with multiplicity $v-1$). The eigenvalue of $M \otimes J_{b/c}$ are the products of the eigenvalue of M with the eigenvalues (b/c and 0) of $J_{b/c}$. Accordingly the eigenvalues of M are

$$0,\ -(k-x)\frac{c}{b},\ \lambda v\frac{c}{b},$$

(with multiplicities $v-1-b+c$, $b-v$ and 1 respectively). □

The above proof shows that $v-1+c-b \geq 0$. It follows that if classes of $\mathcal{D}$ form a partition into 1-designs, the 1-designs are dual 2-designs.

Chapter 10

t-Designs

A t-design is a block-regular incidence structure such that each subset of t points is incident with the same number of blocks. We denote the number of blocks that contain a given set of i points by λ_i (if it is defined). With this convention we have $\lambda_0 = b$ and $\lambda_1 = r$. A t-design where $\lambda_t = 1$ is called a *Steiner system*.

Although t-designs where $t \geq 3$ are not easy to find, they include some very interesting structures.

Hadamard matrices provide the most accessible class of examples. Suppose H is a $v \times v$ Hadamard matrix, with the first row equal to $\mathbf{1}$. Then each row other than the first has an equal number of 1's and -1's, and hence determines two complementary subsets of $\{1, \ldots, v\}$ of size $v/2$. The combined set of $2v - 2$ blocks forms a 3-$(v, v/2, \lambda_3)$ design. Counting pairs consisting of an ordered triple of distinct points and a block which contains them, we find that

$$(2v - 2)v(v - 2)(v - 4)/8 = v(v - 1)(v - 2)\lambda_3$$

whence $\lambda_3 = \frac{1}{4}(v - 4)$. The Möbius planes, which we met earlier, are 3-$(q^2 + 1, q + 1, 1)$ designs.

10.1 Basics

If $\mathcal{D}$ is an incidence structure with point set V and $t \geq 1$, we can form the incidence structure $\mathcal{D}_{\{t\}}$ whose points are the t-subsets of V and whose blocks are the blocks of $\mathcal{D}$, where are t-subset is incident with the blocks

that contain it. (If we are greedy and choose t too large, our incidence structure will not be very interesting.) We call this a *derived structure*.

A remark on λ: if we refer to a t-(v, k, λ) design, then λ denotes λ_t. We also recall that $\lambda_0 = b$ and $\lambda_1 = r$.

10.1.1 Lemma. *Suppose $\mathcal{D}$ is a t-design and $t \geq 2$. If x is a point in $\mathcal{D}$, the points not equal to x and the blocks on x form a $(t-1)$-design. Also the points not equal to x and the blocks not incident with x form a $(t-1)$-design.* □

The blocks on x gives form the derived design of $\mathcal{D}$ relative to x. The blocks off x form the complement to the derived design (relative to x) of the complement of $\mathcal{D}$. We note one consequence of the lemma.

10.1.2 Corollary. *If $\mathcal{D}$ is a 3-(v, k, λ) design, then $b \geq 2v - 2$.*

Proof. Suppose x is a point. The blocks on x form a 2-design on $v-1$ points and so there must be at least $v-1$ blocks on x. Since the blocks off x form a 2-design on $v-1$ points, there must be at least $v-1$ blocks off x. Hence $\mathcal{D}$ has at least $2v-2$ blocks. □

This bound is tight for the Hadamard 3-designs.

If $\mathcal{D}'$ is the derived design of $\mathcal{D}$, we say that $\mathcal{D}$ is an *extension* of $\mathcal{D}'$. We can always begin the construction of an extension of $\mathcal{D}$ by adjoining a new point (traditionally ∞) and adding it to each block of $\mathcal{D}$. This gives us $\lambda_0(\mathcal{D})$ blocks for $\mathcal{D}'$, the remaining blocks are $k+1$ subsets of the point set of $\mathcal{D}$. If $i > 0$, we have

$$\lambda_i(\mathcal{D}') = \lambda_{i-1}(\mathcal{D}).$$

10.1.3 Lemma. *The number of blocks in an extension of a t-(v, k, λ) design is $b(v+1)/(k+1)$.*

Proof. The extension will have parameters $(t+1)$-$(v+1, k+1, \lambda)$, where $\lambda = \lambda_t(\mathcal{D})$. We compute $\lambda_0(\mathcal{D}')$. We have

$$(v+1)\lambda_1(\mathcal{D}') = \lambda_0(\mathcal{D}')(k+1)$$

and thus, if $b = \lambda_0(\mathcal{D})$, we have

$$\lambda_0(\mathcal{D}') = \frac{(v+1)b}{k+1}.$$ □

Exercise: If a projective plane of order n admits an extension, then $n+2$ divides 12 and so $n \in \{2, 4, 10\}$.

An extension of the projective plane of order four has $21 \times 22/6 = 77$ blocks and has parameters 3-$(22, 6, 1)$. The number of blocks in an extension of this would be

$$77 \times 23/7 = 253$$

and these blocks would form a 4-$(23, 7, 1)$ design. The number of blocks in an extension of this would be

$$253 \times 24/8 = 759$$

and we would get a 5-$(24, 8, 1)$. In fact these designs exist, and are known as the *Witt designs*.

Similarly, starting with the affine plane of order three, there are designs with parameters

$$2-(9,3,1),\ 3-(10,4,1),\ 4-(11,5,1),\ 5-(12,6,1)$$

There are also known as Witt designs. The second design here is a Möbius plane.

10.2 Möbius Planes

We introduce our first class of t-designs where $t > 2$. We start with an infinite construction. Let V be the unit sphere in $\mathbb{R}^3$. Define a *circle* to be a subset of C of V such $|C| > 1$ and C is the intersection of V with a plane in $\mathbb{R}^3$ (not necessarily containing the origin). Then any three distinct points on the sphere lie in a unique circle, and we have a 3-design.

The problem is we want finite examples. Let $\mathbb{F}$ be a field, and let $\mathbb{E}$ be a quadratic extension of $\mathbb{F}$. In other words $\mathbb{E}$ is a vector space of dimension two over $\mathbb{F}$. Let σ be an automorphism of $\mathbb{E}$ with order two that fixes each element of $\mathbb{F}$. One examples comes from $\mathbb{R}$ and $\mathbb{C}$, with complex conjugation as σ. A second arises if we take $\mathbb{F}$ to be a finite field of order q and define

$$x^\sigma := x^q.$$

Then

$$(x^\sigma)^\sigma = x^{q^2} = x$$

and so σ has order two. If q is a power of the prime p, then

$$(x+y)^p = x^p + y^p$$

from which it follows that σ is an automorphism. Since the multiplicative group of $\mathbb{F}$ has order q, it follows that σ fixes each element of $\mathbb{F}$.

Now we work over $\mathbb{E}$. If

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

is an invertible matrix over $\mathbb{E}$, define the map τ_A by

$$\tau_A(x) = \frac{ax+b}{cx+d}.$$

We view this as a map from $\mathbb{E} \cup \infty$ to itself, where ∞ satisfies all the obvious rules that you were not permitted to use in Calculus. In particular

$$\tau_A(\infty) = \frac{a}{c}.$$

(Since A is invertible, a and c are not both zero and therefore $\frac{a}{c}$ is a well-defined element of $\mathbb{E} \cup \infty$.) We note that if B is a non-zero scalar multiple of A, then $\tau_A = \tau_B$.

The set of maps τ_A, where A runs over the invertible 2×2 matrices with elements from $\mathbb{E}$ is a group: you may check that

$$\tau_A \circ \tau_B = \tau_{AB}$$

and the multiplication is associative because it is composition of functions. It is denoted by $PGL(2, \mathbb{E})$. It has a subgroup consisting of the maps τ_A where the entries of A are in $\mathbb{F}$. This is denoted by $PGL(2, \mathbb{F})$ and it fixes $\mathbb{F}$ as a set. Since $PGFL(2, \mathbb{F})$ is isomorphic to the group of 2×2 matrices modulo its center, its order is

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

The index of $PGL(2, \mathbb{F})$ in $PGL(2, \mathbb{E})$ is

$$\frac{q^6 - q^2}{q^3 - q} = q(q^2+1).$$

10.2.1 Theorem. *Let $\mathbb{F}$ be a a field of order q and let $\mathbb{E}$ be a quadratic extension of $\mathbb{F}$. Then the images of $\mathbb{F} \cup \infty$ under the action of $PGL(2, \mathbb{E})$ form a 3-$(q^2+1, q+1, 1)$ design.*

Proof. Exercise.

10.3 Doubling 3-$(v, 4, 1)$ Designs

There is a Steiner triple system on v points if and only if $v \equiv 1, 3$ modulo six. The number of blocks in an extension to a 3-$(v+1, 4, 1)$ design is

$$\frac{v+1}{4}\frac{v(v-1)}{6} = \frac{1}{24}(v+1)v(v-1)$$

and since this expression is an integer when $v \equiv 1, 3$ modulo six, the existence of an extension it not ruled out by any simple divisibility condition. It is a major result that a 3-$(v, 4, 1)$ exists whenever $v \equiv 2, 4$ modulo six. Here we present a doubling construction which will provide us with infinitely many 3-designs.

10.3.1 Lemma. *Let $\mathcal{D}$ be a 3-$(v, 4, 1)$ design with point set V and V_1 be a susbset of V of size v_1. If some subset of the blocks of $\mathcal{D}$ form a 3-design with point set V_1, then $v \geq 2v_1$. If equality holds, then any block of $\mathcal{D}$ meets V_1 in an even number of points, and the blocks disjoint from V_1 also form a 3-design.*

Proof. Let $\mathcal{D}_1$ denote the 3-design with point set V_1 V_2 denote $V \setminus V_1$.

We derive the inequality. Choose points y in V_1 and $x \in V_2$. Then for each point z in $V_1 \setminus y$, there is a unique block α that contains x, y and z. Snice each triple of points in V_1 lies in a unique block, the fourth element of which also lies in V_1, we see that α contains exactly two points from V_1. This supplies us with a injection from $V_1 \setminus y$ to the points in $V_2 \setminus x$, and consequently $v \geq 2v_1$.

Assume now that $v = 2v_1$. We prove that any block meets V_1 in an even number of points. If α is block that contains three points of V_1, then the fourth point of α lies in V_1. So the problem is to show that no block of $\mathcal{D}$ contains exactly one point from V_1. Suppose $y \in V_1$. Then exactly $(v-1)(v-2)/6$ blocks of $\mathcal{D}$ lie on y and also exactly

$$\frac{(v_1-1)(v_1-2)}{6} = \frac{(v-2)(v-4)}{24}$$

blocks from $\mathcal{D}_1$ lie in y. So the number of blocks on y not in $\mathcal{D}_1$ is

$$\frac{(v-1)(v-2)}{6} - \frac{(v-2)(v-4)}{24} = \frac{4v^2 - 12v + 8 - v^2 + 6v - 8}{24} = \frac{v^2 - 2v}{8}.$$

As before, for each point z in $V_1 \setminus y$ and each point x in V_2, there is a unique block containing x, y and z and meeting V_1 in eactly two points. Therefore the number of blocks that contain y and exactly one other point from V_1 is

$$\frac{1}{2}\left(\frac{v}{2}-1\right)\frac{v}{2} = \frac{v^2-2v}{8}.$$

Therefore no block of $\mathcal{D}$ contains exactly one point from V_1.

Finally we show that the blocks of $\mathcal{D}$ contained in V_2 form a 3-design. Choose three distinct points in V_2. These lie in a unique block β. Since β can therefore contain at most one point from V_1, and since it cannot contain exactly one point from V_1, we see that $\beta \subseteq V_2$. □

This lemma provides us with an obvious strategy to constructing 3-designs with block size four. Take two such designs on disjoint sets V_1 and V_2 of the same size, and then blocks that contain exactly two points from each set. Here we can use 1-factors and 1-factorizations.

Suppose we have a 3-design as in the lemma with $v = 2v_1$. If y, z are points in V_1, then the blocks in $\mathcal{D}$ that are not subsets of V_1 form a partition of V_2 into disjoint pairs. Thus they are a 1-factor in the complete graph on V_2. If z' is a point inV_1 distinct from y and z, we get a second 1-factor which is disjoint from the first. So we find that for each point y in V_1, there is a 1-factorization of the complete graph on V_2. Similarly each pair of points from V_2 is associated with a 1-factor of the complete graph on V_1, and the 1-factors associated with the pairs that contain a given point of V_2 form a 1-factorization. The points, pairs, 1-factors and 1-factorizations form an incidence structure (with symmetrized inclusion as the incidence relation), and the incidence atructure on V_2 is dual to that on V_1.

With the above ideas as a basis, we arrive at a doubling construction. Let $\mathcal{D}$ be a 3-$(v, 4, 1)$ design with point set V. We construct a 3-$(2v, 4, 1)$ design on two copies of V. Let $F_1, \ldots, F_{v-1}$ and $G_1, \ldots, G_{v-1}$ be two 1-factorizations of K_v. On each copy of V we install a copy of $\mathcal{D}$. View the 1-factors F_i as sets of pairs from the first copy of V and the 1-factors G_i as sets of pairs from the second copy. The number of blocks we need to add to complete our construction is

$$\begin{aligned}\frac{2v(2v-1)(2v-2)}{24} - 2\frac{v(v-1)(v-2)}{24} &= 2v(v-1)\frac{4v-2-v+2}{24}\\ &= \frac{v^2(v-1)}{4}.\end{aligned}$$

To get these, for each pair of 1-factors (F_i, G_i), take all $v^2/4$ of the 4-tuples formed from an edge of F_i and edge of G_i. We need to check that any triple of points lies in exactly one block, but we leave this as an exercise.

We could take the two 1-factorizations to be the same.

10.4 Incidence Matrices

Let $\mathcal{D}$ be a design on point set V of size v. We use N_t to denote the incidence matrix of t-subsets of V versus blocks of $\mathcal{D}$, so $(N_t)_{\alpha,\beta} = 1$ if and only if the t-subset α is contained in the k-subset β.

We use $W_{t,k}(v)$ to denote the 01-matrix with rows indexed by t-subsets of V, columns by k-subsets and with (α, β)-entry equal to 1 if and only if the t-subset α is contained in the k-subset β. When v is clear from the context or irrelevant, we write $W_{t,k}$ for $W_{t,k}(v)$. Note that $W_{t,k}$ is N_t for the complete design with block size k.

10.4.1 Lemma. *If $k\ell$ then*

$$W_{t,k}W_{k,\ell} = \binom{\ell - t}{k - t} W_{t,\ell}$$

Proof. If $|\alpha| = t$ and $|\beta| = \ell$, then the (α, β)-entry of the product is the number of k-subsets γ such that $\alpha \subseteq \gamma\beta$. □

If A and B are matrices and the product AB is defined then the row space of AB is containedin the row space of B. An immediate consequence of this and the above lemma is that each row of $W_{t,\ell}$ lies in the row space of $W_{k,\ell}$.

10.4.2 Lemma. *If $\mathcal{D}$ is a t-design and $s \leq t$, then $\mathcal{D}$ is an s-design.*

Proof. First we note that $\mathcal{D}$ is an s-design if and only if the row sums of N_s are all equal, that is, if $N_s\mathbf{1} = c\mathbf{1}$ for some c. If $\mathcal{D}$ is a t-design, then $N_t\mathbf{1} = \lambda_t\mathbf{1}$ and it will be enough to show that each row of N_s lies in row(N_t).

If σ and β are respectively s-subsets and k-subsets of V, then $(W_{s,t}N_t)_{\sigma,\beta}$ is equal to the number of t-subsets of β that contain σ—hence it is $\binom{k-s}{t-s}$ if $\sigma \subseteq \beta$ and zero otherwise. Consequently

$$W_{s,t}N_t = \binom{k - s}{t - s} N_s. \qquad \square$$

10.5 Complements and Incidence Matrices

Let $\overline{W}_{i,j}$ denote the 01-matrix with rows indexed by the i-subsets of $\{1, \ldots, v\}$, columns by the j-subsets and with (α, β)-entry equal to 1 if $\alpha \cap \beta = \emptyset$. Both parts of the next result are straightforward to prove, and are left as exercises.

10.5.1 Lemma. *We have*

(a) $\overline{W}_{i,k} W_{t,k}^T = \binom{v-t-i}{k-t} \overline{W}_{i,t}$,

(b) $W_{i,k} \overline{W}_{t,k}^T = \binom{v-t-i}{k-i} \overline{W}_{i,t}$. □

10.5.2 Lemma. *We have*

(a) $\overline{W}_{t,k} = \sum_i (-1)^i W_{i,t}^T W_{i,k}$,

(b) $W_{t,k} = \sum_i (-1)^i W_{i,t}^T \overline{W}_{i,k}$.

Proof. We prove (a) and leave (b) as an exercise. Suppose that α is a t-subset of V and β a k-subset. The $\alpha\beta$-entry of $\overline{W}_{t,k}$ is 1 or 0 according as β is contained in the complement of α, or not. The $\alpha\beta$-entry of $W_{i,t}^T W_{i,k}$ is

$$\binom{|\alpha \cap \beta|}{i},$$

while the corresponding entry of the sum in (a) is

$$\sum_i (-1)^i \binom{|\alpha \cap \beta|}{i} = \begin{cases} 1, & \text{if } \alpha \cap \beta = \emptyset; \\ 0, & \text{otherwise.} \end{cases}$$

This completes the proof. □

10.5.3 Lemma. *If $t \leq k \leq v - t$, the matrices $W_{t,k}$ and $\overline{W}_{t,k}$ have the same row space over the rationals.*

Proof. We have $W_{i,k} = \binom{k-i}{t-i}^{-1} W_{i,t} W_{t,k}$ and thus Lemma 10.5.2(a) implies that

$$\overline{W}_{t,k} = \left(\sum_i (-1)^i \binom{k-i}{t-i}^{-1} W_{i,t}^T W_{i,t} \right) W_{t,k}.$$

Therefore each row of $\overline{W}_{t,k}$ is a linear combination of rows of $W_{t,k}$. It is easy to verify that

$$W_{i,t}\overline{W}_{t,k} = \binom{v-k-i}{t-i}\overline{W}_{i,k},$$

whence Lemma 10.5.2(b) implies that

$$W_{t,k} = \left(\sum_i (-1)^i \binom{v-k-i}{t-i}^{-1} W_{i,t}^T W_{i,t}\right)\overline{W}_{t,k}.$$

Consequently each row of $W_{t,k}$ is a linear combination of rows of $\overline{W}_{t,k}$. □

The following consequence of the above lemma underlies many of the applications of linear algebra to combinatorics.

10.5.4 Theorem. *The rank of $W_{t,k}$ is the minimum of its number of rows and its number of columns,*

Proof. Assume $t \le k \le v-t$. We first consider the case where $v = t+k$. Then $W_{t,v-t}$ and $\overline{W}_{t,k}$ are square of the same order. As $\overline{W}_{t,v-t}$ is a permutation matrix, it is invertible. Since $\overline{W}_{t,v-t}$ and $W_{t,v-t}$ have the same row space, they have the same rank and thus $W_{t,v-t}$ is invertible.

Now if $t \le h \le v-t$ then

$$W_{t,h}W_{h,v-t} = \binom{v-2t}{h-t}W_{t,v-t}.$$

Since the matrix on the right of this equation is invertible, it follows that the rows of $W_{t,h}$ are linearly independent. □

10.5.5 Lemma. *A design and its complement have the same strength.*

Proof. From the proof of the Lemma 10.5.3 we know that there are matrices, G and H say, such that

$$\overline{W}_{t,k} = GW_{t,k}, \qquad W_{t,k} = H\overline{W}_{t,k}.$$

Further the rows sums of G are constant, as are the row sums of H. Hence if $W_{t,k}x = \lambda\mathbf{1}$ then $\overline{W}_{t,k}x = \lambda G\mathbf{1} = c\mathbf{1}$ for some constant c. Since $W_{t,v-k}$ is got from $\overline{W}_{t,k}$ by permuting its columns, it follows that the complement of a design with strength t has strength at least t. The lemma follows immediately. □

10.6 Extending Fisher's Inequality

Let $\varphi(t, v)$ denote the minimum number of blocks in a t-design on v points. From Lemma 10.1.1 we have the inequality

$$\varphi(t+1, v) \geq 2\varphi(t, v-1).$$

As $\varphi(2, v) = v$, we see that a 3-design on v must have at least $2v-2$ blocks. This shows that our inequality is tight for one value of t, at least (and that Hadamard matrices again give rise to an exceptional class of designs).

The bound we have just found is not tight for $t > 3$. However we have an important extension of Fisher's inequality, due to Ray-Chaudhuri and Wilson.

10.6.1 Theorem. *If $\mathcal{D}$ is a t-design with b blocks and v points, then*

$$b \geq \binom{v}{\lfloor \frac{t}{2} \rfloor}.$$

Proof. If N_i is the incidence matrices for i-subsets versus blocks and $2i \leq t$, then the (α, β)-entry of $N_i N_i^T$ is equal to $\lambda_{|\alpha \cup \beta|}$. As

$$\frac{\lambda_i}{\lambda_j} = \frac{\binom{v-i}{k-i}}{\binom{v-j}{k-j}}$$

it follows that

$$\frac{1}{\lambda_i} N_i N_i^T = \frac{1}{\binom{v-i}{k-i}} W_{i,k} W_{i,k}^T.$$

Since the rows of $W_{i,k}$ are linearly independent, the rows of N_i must be linearly independent too, and this yields the bound. □

If equality holds in this bound then we say that $\mathcal{D}$ is *tight*. Symmetric designs are thus tight 2-designs.

For a tight 4-design with $\lambda_4 = 1$, we have

$$\binom{v}{2} k(k-1)(k-2)(k-3) = v(v-1)(v-2)(v-3)$$

and so

$$\binom{k}{2}\binom{k-2}{2} = \binom{v-2}{2}.$$

10.7 Intersection Triangles

We work with t-designs on the set $\{1, \ldots, v\}$. If i and j are non-negative integers and $\alpha = \{a_1, \ldots, a_{i+j}\}$ is a sequence of distinct points, then $\lambda_{i,j}(\mathcal{D})$ denotes the number of blocks β of $\mathcal{D}$ such that

$$\beta \cap \{a_1, \ldots, a_{i+j}\} = \{a_1, \ldots, a_i\}.$$

We note that $\lambda_{i,0} = \lambda_i(\mathcal{D})$ and, if $\overline{\mathcal{D}}$ denotes the complement of $\mathcal{D}$, then $\lambda_{0,i}(\mathcal{D}) = \lambda_i(\overline{\mathcal{D}})$. We also have

$$\lambda_{i,j} = \lambda_{i+1,j} + \lambda_{i,j+1}. \tag{10.7.1}$$

This leads to a version of Pascal's triangle. If $t = 2$, then each entry in the triangle

$$\begin{array}{ccccc} & & b & & \\ & r & & b-r & \\ \lambda & & r-\lambda & & b-2r+\lambda \end{array}$$

is the sum of the two entries immediately below it, and the j-th entry in row i is equal to $\lambda_{i-j,j}$.

Given the recurrence in (10.7.1) an easy induction argument yields the following:

10.7.1 Lemma. *If $\mathcal{D}$ is a t-design and $i + j \leq t$, then $\lambda_{i,j}$ is determined by the parameters of $\mathcal{D}$.* □

One corollary of this lemma is that the complement of a t-design is a t-design. You may show that

$$\lambda_s(\overline{\mathcal{D}}) = \lambda_{0,s} = \sum_{i=0}^{s} (-1)^i \binom{s}{i} \lambda_{i,0}.$$

In some interesting cases we can compute $\lambda_{i,j}$ even when $i + j > t$. Suppose $a_1, \ldots, a_k$ is a Steiner system with block size k and let $a_1, \ldots, a_k$ be the points in the block α. Then we can compute $k+1$ rows of the intersection triangle, because the number of blocks that intersect α in $a_1, \ldots, a_j$ is i when $j \geq t$. By way of example we offer the intersection triangle for a 4-$(23, 7, 1)$ design

						253								
					77			176						
				21			56			120				
			5			16			40			80		
		1			4			12			28			52
	1			0			4			8			20			32	
1			0			0			4			4			16			16
1		0			0			0			4			0			16			0

From the last row of this table we deduce that for a design with these parameters, any two distinct blocks meet in 1 or 3 points. Hence it is a quasisymmetric design.

10.8 Polynomials

If M is a matrix and $p(x)$ is a polynomial, we define $p \circ M$ to be the matrix with the same order as M, with

$$(p \circ M)_{i,j} := p(M_{i,j}).$$

10.8.1 Lemma. *Let $\mathcal{D}$ be a t-design with incidence matrix N and suppose $p(x)$ is a polynomial of degree d, where $d \leq k$. If $M_r := N_r^T N_r$, then $p \circ M_1$ is a linear combination of the matrices $M_0, \ldots, M_d$.*

Proof. There are scalars $c_{r,s}$ such that

$$x^r = \sum_s c_{r,s} \binom{x}{s}$$

and

$$\left(\binom{x}{s} \circ M_1 \right)_{\alpha,\beta} = \binom{|\alpha \cap \beta|}{s} = (M_s)_{\alpha,\beta}.$$

10.8.2 Corollary. *If* $\deg(p) = s$ *and* $s \leq k$, *then* $\mathrm{rk}(p \circ M_1) \leq \binom{v}{s}$.

Proof. Assume $i \leq j$. Since

$$W_{i,j}W_{j,\ell} = \binom{\ell - i}{j - i} W_{i,\ell},$$

we have

$$W_{i,j}N_j = N_i$$

and consequently $\mathrm{row}(N_j)$ is contained in $\mathrm{row}(N_i)$. Since $\mathrm{row}(N_i) = \mathrm{row}(N_i^T N_i)$, it follows that $\mathrm{row}(M_i) \leq \mathrm{row}(M_j)$. Therefore $\mathrm{row}(p \circ M_1)$ is contained in $\mathrm{row}(M_s)$, which has dimension $\binom{v}{s}$. □

The following result is also due to Ray-Chaudhuri and Wilson.

10.8.3 Theorem. *If* $\mathcal{D}$ *is a simple design on* v *points with degree* s, *then* $|\mathcal{B}| \leq \binom{v}{s}$.

Proof. Let Δ be the degree set of $\mathcal{D}$ and let p be the monic polynomial of degree s with the elements of Δ as its zeros. Then

$$p \circ M_1 = p(k)I$$

and accordingly

$$b = \mathrm{rk}(p(k))I = \mathrm{rk}(p \circ M_1) \leq \binom{v}{s}.$$ □

10.9 Gegenbauer Polynomials

Let G_s be the matrix representing orthogonal projection onto the columns of $W_{s,k}^T$. Thus $G_s = W_{s,k}^T (W_{s,k} W_{s,k}^T)^{-1} W_{s,k}$; we derive a more explicit expression for it.

10.9.1 Lemma. *We have*

$$G_s = \sum_{i=0}^{s} (-1)^i \frac{\binom{k-i}{s-i}}{\binom{v-s-i}{k-s}} W_{i,k}^T \overline{W}_{i,k}.$$

Proof. Take G_s to be the matrix just defined. We have to prove that it is the required projection. We know that $W_{i,k}$ and $\overline{W}_{i,k}$ have the same row space, and so it follows that the column space of G_s is contained in the column space of $W_{s,k}$. Therefore $G_s x$ lies in the column space of $W_{s,k}^T$, for any vector x, and so it will suffice to prove that $G_s W_{s,k}^T = W_{s,k}^T$. By Lemma 10.5.2(a) we find that

$$G_s W_{s,k}^T = \sum_i (-1)^i \binom{k-i}{s-i} W_{i,k}^T \overline{W}_{i,s}. \tag{10.9.1}$$

Since

$$\binom{k-i}{s-i} W_{i,k}^T = W_{s,k}^T W_{i,s}^T$$

the right side of the previous equation is equal to

$$W_{s,k}^T \sum_i (-1)^i W_{i,s}^T W_{i,s}.$$

By Lemma 10.5.2(b), the sum here is equal to $\overline{W}_{t,t} = I$, and thus the lemma follows. □

10.9.2 Corollary. *If α and β are k-subsets of a v-set then*

$$(G_s)_{\alpha,\beta} = \sum_{i \geq 0} (-1)^i \frac{\binom{k-i}{s-i}\binom{k-|\alpha\cap\beta|}{i}}{\binom{v-s-i}{k-s}}.$$ □

This corollary implies that the (α, β)-entry of G_s is a polynomial in $|\alpha \cap \beta|$ with degree at most s. We define the *Gegenbauer polynomial* g_s by

$$g_s(x) = \binom{v}{k} \sum_{i \geq 0} (-1)^i \frac{\binom{k-i}{s-i}\binom{k-x}{i}}{\binom{v-s-i}{k-s}}.$$

Note that $g_s(k) = \binom{v}{s}$.

10.10 A Positive Semidefinite Matrix

We have seen already that a t-design on v points has at least

$$\binom{v}{\lfloor \frac{t}{2} \rfloor}$$

blocks. We now prove a result which provides even more information (at somewhat greater cost). Let Ω denote the set of all k-subsets of the v-set V. If $\Phi \subseteq \Omega$, let $G_s(\Phi)$ denote the principal submatrix of G_s with rows and columns indexed by the elements of Φ. What we need is essentially an upper bound on the largest eigenvalue of $G_s(\Phi)$, which we derive indirectly.

10.10.1 Lemma. *If Φ is a t-design and $2r \leq t$ then $|\Phi|^{-1}\binom{v}{k}G_r(\Phi)$ is idempotent.*

Proof. First recall that

$$G_s = W_{s,k}^T(W_{s,k}W_{s,k}^T)^{-1}W_{s,k}$$

and therefore

$$G_s(\Phi) = N_s^T(W_{s,k}W_{s,k}^T)^{-1}N_s.$$

If Φ is a t-design and $2s \leq t$, then

$$|\Phi|^{-1}N_sN_s^T = \binom{v}{k}^{-1}W_{s,k}W_{s,k}^T.$$

whence

$$G_s(\Phi) = \frac{|\Phi|}{\binom{v}{k}}N_s^T(N_sN_s^T)^{-1}N_s.$$

Since $N_s^T(N_sN_s^T)^{-1}N_s$ is the matrix that represents orthogonal projection onto $\mathrm{col}(N^T)$, we conclude that $|\Phi|^{-1}\binom{v}{k}G_s(\Phi)$ is idempotent. □

As the eigenvalues of an idempotent matrix are all equal to 0 or 1, the following follows immediately.

10.10.2 Corollary. *If Φ is a subset of Ω with strength at least $2r$, then the matrix $|\Phi|I - \binom{v}{k}G_r(\Phi)$ is positive semidefinite.* □

If a matrix is positive semi-definite then any principal submatrix of it is also positive semi-definite. Hence its diagonal entries are all non-negative. What are the diagonal entries of $|\Phi|I - \binom{v}{k}G_r(\Phi)$? The diagonal entries of G_r are all equal to $\binom{v}{r}/\binom{v}{k}$ (why?), whence we see that if Φ has strength $2r$ then

$$|\Phi| \geq \binom{v}{r}.$$

We have proved this already, as Table **??**. Fortunately we can say more.

10.10.3 Theorem. *Let Φ be a subset of Ω with strength at least $2r$. If α and β are distinct elements of Φ with $|\alpha \cap \beta| = i$ then*

$$|\Phi| \geq \binom{v}{r} + |g_r(i)|.$$

Proof. Assume $b = |\Phi|$ and suppose that α and β are elements of Φ such that $|\alpha \cap \beta| = i$. Then $|\Phi| I - \binom{v}{k} G_r(\Phi)$ has a 2×2 submatrix equal to

$$\begin{pmatrix} b - \binom{v}{r} & -g_r(i) \\ -g_r(i) & b - \binom{v}{r} \end{pmatrix}.$$

As this submatrix must be positive semi-definite, its determinant is non-negative. Hence we find that

$$\left(b - \binom{v}{r} - g_r(i)\right)\left(b - \binom{v}{r} + g_r(i)\right) \geq 0,$$

which proves the theorem. □

10.10.4 Theorem. *Let Φ be a subset of Ω with strength at least $2r$. Then $|\Phi| \geq \binom{v}{r}$ and, if equality holds, the degree set of Φ is the set of zeros of g_r.*

Proof. Table **??** shows that $|\Phi| \geq \binom{v}{r}$ and that, if equality holds, $g_r(i) = 0$ for any i in the degree set of Φ. This shows that the degree of Φ is at most r. If it is less than r then $|\Phi| < \binom{v}{r}$, by Table **??**, and therefore its degree is r. □

One consequence of the last theorem is that the degree set of a tight design is determined by v, k and r. Further, most polynomials g_r do not have r integer zeros. Thus we obtain strong restrictions on the parameters of a tight design.

10.11 Polynomial Spaces: Functions

We aim to generalize the calculations related to the inequalities of Ray-Chaudhuri and Wilson. The idea is to view a design as a subset of a larger set.

We start with a 'universal' set Ω. Relevant examples are the set of all k-subsets of a set of size v, the elements of the symmetric group $\mathrm{Sym}(n)$, the unit vectors in $\mathbb{R}^d$. The set Ω comes with an injective embedding into a real (or complex) inner product vector space. Thus for k-sets we use the map that takes a k-set to its characteristic function, in $\mathbb{R}^v$. For the unit vectors, we use the vectors themselves. We will usually be sloppy and identify Ω with its image in the vector space. We view an orthogonal array $OA(k, n)$ as a subset of the set of all functions from $\{1, \ldots, k\}$ to $\{1, \ldots, \}$, and we denote this set of functions by $H(k, n)$.

If $\alpha \in \Omega$ then the coordinate maps

$$\alpha \mapsto \langle e_i, \alpha \rangle$$

are functions on Ω, we will view them as linear functions on Ω. We define a sequence of space of functions $\mathrm{Pol}(\Omega, r)$ by setting $\mathrm{Pol}(\Omega, 0)$ equal to the space of constant functions, setting $\mathrm{Pol}(\Omega, 1)$ equal to the space spanned by the constant and linear functions and then inductively defining $\mathrm{Pol}(\Omega, r)$ by setting

$$\mathrm{Pol}(\Omega, r+1) = \mathrm{Pol}(\Omega, 1) \cdot \mathrm{Pol}(\Omega, r+1)$$

when $r \geq 1$. The union all the spaces $\mathrm{Pol}(\Omega, r)$ will be denoted by $\mathrm{Pol}(\Omega)$. We use 1 to denote the constant function taking the value 1 on Ω.

When Ω is the set of all k-subsets of a v-set, we denote it by $J(v, k)$. We see that the rows of the Wilson matrix $W_{1,k}(v)$ are the coordinate functions on $J(v, k)$, and rows of $W_{t,k}(v)$ lie in $\mathrm{Pol}(\Omega, t)$.

We use $H(n, d)$ to denote the set of all vectors of length n with entries from some set D of size d. We can view both the rows of an orthogonal array and permutations from $\mathrm{Sym}(n)$ as elements of $H(n, d)$ (with $|D| = n$ in the case of the symmetric group). In practice we prefer a second approach. We represent the elements of an orthogonal array $OA(k, n)$ by vectors of length kn—the idea is to represent i in $\{1, \ldots, N\}$ by the i-th standard basis vector e_i, which expands each row of an $OA(k, n)$ to a 01-vector in $\mathbb{R}^{kn}$. The coordinate functions arise by taking inner products with the standard basis vectors in $\mathbb{R}^{kn}$.

Because we have an embedding of Ω in an inner product space, if $\alpha \in \Omega$, we can define a function z_α on Ω by

$$z_\alpha(\beta) = \langle \alpha, \beta \rangle$$

Clearly z_α is linear and, in all cases of interest to us, the span of the functions z_α for α in Ω will be $\mathrm{Pol}(\Omega, 1)$.

10.11.1 Lemma. *If $\Omega = J(v,k)$, then $\mathrm{Pol}(\Omega, i)$ is equal to the row space of N_i.*

Proof. Each row of $W_{i,k}$ is a Schur product of i distinct rows of $W_{1,k}$, and each product of i rows of $W_{1,k}$ lies in the row space of $W_{i,k}$. □

10.12 Polynomial Spaces: Averaging, Designs

So far we have defined a space of functions/polynomials. The next step is to add an inner product on $\mathrm{Pol}(\Omega)$. However we insist that that this inner product have the property that, for any two functions f and g in $\mathrm{Pol}(\Omega)$,

$$\langle f, g\rangle = \langle 1, fg\rangle.$$

We insist further that, if f is non-negative then $\langle 1, f\rangle \geq 0$, and that if $f \geq 0$ and $\langle 1, f\rangle = 0$ then $f = 0$.

There is a way to construct such inner products. Let ν be a linear functional on $\mathrm{Pol}(\Omega)$. Then the map

$$(f, g) \mapsto \nu(fg)$$

is symmetric and bilinear. It is non-degenerate if and only if $\ker \mu$ does not contain an ideal of $\mathrm{Pol}(\Omega)$, and our non-negativity constraint will hold provided that:

(a) $\mu(f) \geq 0$ when $f \geq 0$, and

(b) $\ker \nu$ does not contain a non-negative function.

But this is making something relatively simple look complicated; in practice our function μ will be some sort of average on functions in $\mathrm{Pol}(\Omega)$.

When Ω is finite, we define

$$\langle 1, f\rangle = \mu(f) = \frac{1}{|\Omega|}\sum_{\alpha\in\Omega} f(\alpha)$$

For the unit sphere, $\mu(f)$ is the usual average of f over the sphere. (Note here that there are non-negative functions on the unit sphere with average value 0, but no non-negative polynomials average to zero.

If Φ is a finite subset of Ω, we define a bilinear form on $\mathrm{Pol}(\Omega)$ by

$$\langle f, g\rangle_\Phi = \frac{1}{|\Phi|}\sum_{\alpha\in\Phi} f(\alpha)g(\alpha).$$

(This is not necessarily an inner product, because it may be that f is not zero, but its restriction to Φ will be zero.) It is at least positive semidefinite.

We define a finite subset Φ to be a *t-design* if, for all f in $\mathrm{Pol}(\Omega, t)$, we have

$$\langle 1, f\rangle_\Phi = \langle 1, f\rangle.$$

The *strength* of Φ is the largest value of t for which Φ is a t-design.

10.12.1 Lemma. *A subset Φ of $J(v,k)$ is a t-design in the polynomial space if and only if it is a t-design in the usual sense.*

Proof. The subset Φ is a t-design in the usual sense if and only if, for each row f of $W_{t,k}$, the sum

$$\sum_{\alpha\in\Phi} f(\alpha) = \lambda_t$$

for some constant λ_t and it follows immediately that a t-design in the polynomial space sense is a t-design in the usual sense.

For the converse, if the previous equation holds then summing it over all rows yields

$$\binom{v}{t}\lambda_t = \sum_f \sum_{\alpha\in\Phi} f(\alpha) = \binom{k}{t}|\Phi|$$

and therefore

$$\langle 1, f\rangle_\Phi = \frac{1}{|\Phi|}\lambda_t = \frac{\binom{k}{t}}{\binom{v}{t}} = \frac{\binom{v-t}{k-t}}{\binom{v}{k}} = \langle 1, f\rangle. \qquad \square$$

10.13 Polynomial Spaces: Codes

We required that the vector space in which Ω embeds should be an inner product space, and now we make use of this. We denote the inner product by ρ. If $\Phi \subseteq \Omega$, the *degree set* of Φ is the set of values taken by $\rho(\alpha,\beta)$ as (α,β) ranges over the distinct pairs of elements of Φ, and the *degree* of Φ is the size of its degree set.

We assume that $\rho(\alpha, \alpha)$ is independent of α. One consequence of this is that if $\alpha \neq \beta$, then $\rho(\alpha, \beta) < \rho(\alpha, al)$.

If N_1 is incidence matrix of a design, then the degree of the design is the number of different sizes of the intersections of two distinct blocks; the 2-designs with degree one are precisely the symmetric designs.

If $\alpha \in \Omega$, we define the function ρ_α by

$$\rho_\alpha = \rho(\alpha, \beta).$$

If $p(t)$ is a real polynomial, we define the function p_α by

$$p_\alpha(\beta) = p(\rho_\alpha(\beta));$$

thus p_α is the composition $p \circ \rho_\alpha$. The functions in the space spanned by the functions p_α where $\deg(p) \leq i$ and $\alpha \in \Omega$ is defined to be the space of *zonal polynomials of degree at most* i. In all cases of interest to us, the space of zonal polynomials of degree r is equal $\mathrm{Pol}(\Omega, r)$, but this is not trivial. It is easier to show that zonal polynomials of degree r lie in $\mathrm{Pol}(\Omega, r)$.

10.14 Bounds on Codes and Designs

10.14.1 Theorem. *Suppose Ω is a polynomial space. If the finite subset Φ of Ω is a t-design, then*

$$|\Phi| \geq \dim(\mathrm{Pol}(\Omega, \lfloor t/2 \rfloor)).$$

Proof. Let $f_1, \ldots, f_m$ be an orthonormal basis for $\mathrm{Pol}(\Omega, \lfloor t/2 \rfloor)$. Then for all i and j, the product $f_i f_j$ lies in $\mathrm{Pol}(\Omega, t)$. Hence

$$\delta_{i,j} = \langle f_i, f_j \rangle_\Phi = \langle 1, f_i f_j \rangle_\Phi = \langle 1, f_i f_j \rangle = \langle f_i, f_j \rangle$$

and therefore the restrictions to Φ of the functions f_i are linearly independent functions on Φ. Consequently

$$\dim(\mathrm{Pol}(\Omega, t/2)) = m \leq |\Phi|. \qquad \square$$

10.14.2 Theorem. *Suppose Ω is a polynomial space. If the proper subset Φ of Ω has degree s, then*

$$|\Phi| \leq \dim(\mathrm{Pol}(\Omega, s)).$$

Proof. There is a unique monic polynomial ψ of degree s that vanishes on the degree set of Φ and is equal to 1 on $\rho(\alpha, \alpha)$. The zonal polynomials ψ_α, for α in Φ, are linearly independent, since the intersection of the support of ψ_α with Φ is $\{\alpha\}$. Therefore $|\Phi| \leq \dim(\mathrm{Pol}(\Omega, s))$. □

The polynomial p in the proof the previous theorem is called the *annihilator* of the degree set of Φ.

10.14.3 Theorem. *Let Ω be a polynomial space. If the finite subset Φ has degree s and strength t, then $t \leq 2s$.*

Proof. Let ψ be defined as in the proof of the previous theorem. Let $\delta = \rho(\alpha, \alpha)$ and define $\varphi(t) = (\delta - t)\psi(t)^2$. If $\alpha \in \Phi$, then φ_α is zero on Φ and non-negative on Ω. If $t \geq 2s + 1$ we thus have

$$0 = \langle 1, \varphi_\alpha \rangle_\Phi = \langle 1, \varphi_\alpha \rangle$$

and this implies that φ_α must vanish identically on Ω. Consequently the degree of Ω is at most s and so Ω is finite by the previous theorem. Since $\deg(\psi) = s$,

$$\frac{\psi_a(\delta)}{|\Phi|} = \langle 1, \psi_a \rangle_\Phi = \langle 1, \psi_a \rangle = \frac{\psi_a(\delta)}{\Omega|},$$

whence $|\Phi| = |\Omega|$ and therefore $\Omega = \Phi$. □

Our next result is a form of linear programming bound.

10.14.4 Theorem. *Suppose Φ is a t-design in a polynomial space Ω. If $p \in \mathrm{Pol}(\Omega, t)$ and the restriction of p to Φ is non-negative, then for any α in Φ,*

$$|\Phi| \geq \frac{p(\alpha)}{\langle 1, p \rangle}.$$

Proof. We have

$$\frac{p(\alpha)}{|\Phi|} \leq \langle 1, p \rangle_\Phi = \langle 1, p \rangle.$$ □

Chapter 11

Witt Designs

Our main goal in this section is to construct the Witt designs on 23 and 24 points; these are 4- and 5-designs respectively.

11.1 Codes

A *code* is a subspace of a vector space, along with an implied threat that Hamming distance will be mentioned very soon. The *Hamming distance* between two vectors is the number of coordinate positions in which they differ, and the *weight* of a vector is the number of non-zero entries in it—this is its the Hamming distance from the zero vector. The *length* of a code is the dimension of the vector space that contains it. The vector space V is often $GF(q)^n$ but it could, for example, be the quotient ring $GF(q)[x]/(x^n - 1)$. In the latter case, subspaces which are invariant under multiplication by x are *cyclic* codes.

If V is a vector space and $v \in V$ then we define

$$v^\perp = \{u \in V : v^T u = 0\};$$

this is a subspace of V and, if $v \neq 0$, its codimension in V is 1. If $S \subseteq V$, then

$$S^\perp := \bigcap_{v \in S} v^\perp.$$

If U is a code in V, then $U^\perp$ is the *dual code* of U. If $U \leq U^\perp$ we say that U is *self-orthogonal*, and if $U = U^\perp$ that U is *self-dual*.

Even if $v \neq 0$, we may find that $v \in v^\perp$—for example if $v = \mathbf{1}$ and $\mathbb{F} = \mathbb{Z}_2$ and $\dim(V)$ is even. More generally, the intersection $U \cap U^\perp$ need not be empty. The following useful properties still hold though.

11.1.1 Lemma. *If U is a subspace of the vector space V, then*

(a) $(U^\perp)^\perp = U$.

(b) $\dim(U^\perp) = \dim(V) - \dim(U)$.

If M and N are matrices of over $GF(p)$ and $MN^T = 0$, then $\mathrm{row}(N)$ is a subspace of $\mathrm{row}(M)^\perp$. Hence if the number of columns of M is n,

$$\mathrm{rk}(N) = \dim(\mathrm{row}(N)) \leq \dim(\mathrm{row}(M)^\perp) = n - \mathrm{rk}(M)$$

and therefore $\mathrm{rk}(M) + \mathrm{rk}(N) \leq n$. Thus if $MM^T = 0$ then $2\,\mathrm{rk}(M) \leq n$.

A binary code is *even* if all words have even weight, and it is *doubly even* if all words have weight divisible by four. We leave the proof of the following as an exercise.

11.1.2 Lemma. *Let G be a matrix over $GF(2)$. If each row of G has even weight, then $\mathrm{row}(G)$ is an even code. If $GG^T = 0$ and each row of G is doubly even, then $\mathrm{row}(G)$ is doubly even.* □

Suppose C is the code formed by the row space of a $k \times n$ matrix G over $\mathbb{Z}_2$. Let G^1 be the matrix we get from G by adding an extra coordinate to each code word, where the value of the coordinate is the weight of the code word. The only interesting case is when C contains words of odd weight, because then the extended code is even.

11.2 Codes from Designs

The codes of interest to us will arise usually as the row or column spaces of incidence matrices. We use $\mathrm{rk}_p(N)$ to denote the rank of N, viewed as a matrix over $\mathbb{Z}_p$. We eliminate some uninteresting cases.

11.2.1 Lemma. *Let $\mathcal{D}$ be a 2-design with incidence matrix N and let $n = r - \lambda$ be the order of $\mathcal{D}$. If p is a prime that does not divide n, then $\mathrm{rk}_p(N) \geq v - 1$. If p divides n, then $\mathrm{rk}_p(N) \leq \frac{1}{2}(v+1)$.*

Proof. We have

$$NN^T = nI + \lambda J.$$

If p is a prime that does not divide n, then nI is invertible and $\mathrm{rk}_p(J) = 1$. Therefore $\mathrm{rk}_p(nI + \lambda J) \geq v - 1$.

If $p \mid n$ then

$$NN^T = \lambda J \mod p$$

and therefore $\mathrm{rk}_p(NN^T) \leq 1$. It follows that (prove it!)

$$\mathrm{rk}_p(N) \leq \frac{v+1}{2}. \qquad \square$$

If p does not divide N and $\mathrm{rk}_p(N) = v-1$, you may prove that $\mathrm{col}(N) = \mathbf{1}^\perp$. Note that if $NN^T = \lambda J$ modulo p, then the space spanned by the differences of the rows of N is self-orthogonal.

11.3 Matrix Ranks

We need information about the ranks of matrices over $\mathbb{Z}_p$. The techniques introduced here provide lower bounds.

If $a, b \in \mathbb{Z}$ and the gcd of a and b is d, there are integers x and y such that $xa + yb = d$, then

$$\begin{pmatrix} x & y \\ -b/d & a/d \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} d \\ 0 \end{pmatrix}$$

The 2×2 matix here has determinant one, and so it follows that if α is an integer vector such that the gcd of the elements of α is d and e_1 is the first standard basis vector, there is an integer matrix M with determinant ± 1 such that $M\alpha = de_1$.

An integer matrix with determinant ± 1 is said to be *unimodular*.

11.3.1 Lemma. *If N is a matrix over $\mathbb{Z}$, there are unimodular matrices L and R such that LBR is diagonal.* $\square$

11.3.2 Lemma. *Let M be an $n \times n$ integer matrix and let p be a prime. If p^k divides $\det(M)$, but p^{k+1} does not, then $\mathrm{rk}_p(M) \geq n - k$.*

Proof. Choose unimodular matrices L and R such that $D = LMR$ is diagonal. Note that, up to sign, $\det(D) = \det(M)$; further $\mathrm{rk}_p(D) = \mathrm{rk}_p(M)$. Then at most k diagonal elements of D can be divisible by p, and therefore $\mathrm{rk}_p(D) \geq n - k$. □

11.4 Codes from Projective Planes

11.4.1 Theorem. *Let $\mathcal{P}$ be a projective plane of order n with incidence matrix N. If $n \equiv 2$ modulo four, the extension of the code formed by the rows of N is self-dual and doubly even.*

Proof. Since n is even, k is odd and so

$$\hat{N} = \begin{pmatrix} N & \mathbf{1} \end{pmatrix}$$

is a generator for the extended code. We have

$$\hat{N}\hat{N}^T = NN^T + J = nI + 2J = 0$$

and thus the extended code is self-orthogonal. Since each row of $\hat{N}$ has weight divisible by four, it follows from an exercise that the extended code is doubly even.

To prove that the extended code is self dual, we show that its dimension is $(n^2+n+2)/2$. Since $\mathbf{1} \in \mathrm{col}(N)$ we see that $\mathrm{rk}(\hat{N}) = \mathrm{rk}(N)$. As $\hat{N}\hat{N}^T = 0$ we have that $\mathrm{rk}(\hat{N}) \leq (n^2 + n + 2)/2$.

For the other direction we have

$$\det(NN^T) = (r - \lambda)^{v-1} rk$$

for any 2-design and so for a projective plane,

$$\det(N) = n^{(v-1)2}(n + 1).$$

Since $n \equiv 2$ modulo four, it follows that $2^{(v-1)/2}$ is the largest power of two that divides $\det(N)$, and using Lemma 11.3.2 we deduce that $\mathrm{rk}_2(N) \geq (v + 1)/2$. □

11.5 MacWilliams Theorem

If C is a code we use A_r to denote the number of words of weight r in C. The *weight enumerator* of C is the polynomial

$$W_C(x,y) = \sum_i A_i x^i y^{n-i}.$$

Thus $W_C(1,1) = 2^{\dim(C)}$ for a binary code. it is a surprising fact, due to MacWilliams, that the weight enumerator of a code determines the weight enumerator of its dual. We state the result for binary codes.

11.5.1 Theorem. *If C is a linear code over $GF(q)$,*

$$W_{C^\perp} = \frac{1}{|C|} W_C(-x+y, (q-1)x+y).$$

We give a proof for the case $p = 2$ (this is all that we need, and it is not hard to modify our proof to give the general result). For the proof we need a definition and two lemmas.

Let V be a vector space over $\mathbb{Z}_2$. If f is a function on V, its *Hadamard transform* f^H is defined by

$$f^H(u) = \sum_{v \in V} (-1)^{u^T v} f(v).$$

11.5.2 Lemma. *If C is a binary linear code, then*

$$\sum_{u \in C^\perp} f(u) = \frac{1}{|C|} \sum_{v \in C} f^H(v).$$

Proof. We have

$$\begin{aligned} \sum_{u \in C} f^H(u) &= \sum_{u \in C} \sum_{v \in \mathbb{F}^n} (-1)^{u^T v} f(v) \\ &= \sum_{v \in \mathbb{F}^n} \sum_{u \in C} (-1)^{u^T v}. \end{aligned}$$

If $v \in C^\perp$ then $v^T u = 0$ for all u in C and the inner sum is $|C|$. Otherwise there is a vector c in C such that $v^T c = 1$. Now we can partition C into pairs of the form $(u, u+c)$ where $u \in C$, and $v^T u \neq v^T(u+c)$ for any c in U. It follows that our inner sum is zero, and so the lemma is proved. □

11.5.3 Lemma. *If f is the function on the vector space over $\mathbb{Z}_2$ such that*

$$f(u) = x^{\mathrm{wt}(u)} y^{n-\mathrm{wt}(u)}$$

then

$$f^H(v) = (-x+y)^{\mathrm{wt}(v)}(x+y)^{\mathrm{wt}(v)}.$$

Proof. We have

$$\begin{aligned} F^H(u) &= \sum_{v \in \mathbb{F}^n} \left((-1)^{u_1 v_1 + \cdots + u_n v_n} \prod_{i=1}^{n} (x^{v_i} y^{1-v_i}) \right) \\ &= \sum_{v} \left(\prod_{i=1}^{n} (-1)^{u_i v_i} x^{v_i} y^{1-v_i} \right) \\ &= \prod_{i=1}^{n} ((-1)^{u_i} x + y) \\ &= (-x+y)^{\mathrm{wt}(u)} (x+y)^{n-\mathrm{wt}(u)} \end{aligned}$$

and the lemma follows at once. □

It is now very easy to prove MacWilliams' theorem. If f is defined as in the previous lemma, then

$$\sum_{u \in C^\perp} f(u) = W_{C^\perp}(x, y)$$

and by Lemma 11.5.3,

$$\begin{aligned} \sum_{u \in C^\perp} f(u) &= \frac{1}{|C|} \sum_{v \in C} f^H(v) \\ &= \frac{1}{|C|} \sum_{v \in C} (-x+y)^{\mathrm{wt}(v)} (x+y)^{\mathrm{wt}(v)} \\ &= \frac{1}{|C|} W_C(-x+y, x+y). \end{aligned}$$

11.6 Nonexistence of Some Projective Planes

We use the coding theory we have developed to show the following.

11.6.1 Theorem. *If $n \equiv 6$ modulo 8, there is no projective plane of order n.*

Proof. Assume that $n \equiv 2$ mod four. Then the extended code of the plane is, as we saw, self-dual and doubly even. Since it is doubly even

$$W_C(ix, y) = W_C(x, y)$$

and since it is self-dual

$$W_C(x, y) = W_{C^\perp}(x, y) = \frac{1}{|C|} W_C(-x + y, x + y) = W_C\left(\frac{-x + y}{\sqrt{2}}, \frac{x + y}{\sqrt{2}}\right).$$

So we have two substitutions in the variables x and y that leave the polynomial $W_C(x, y)$ invariant, and therefore the composition of these two substitutions

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -i & 1 \\ i & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

will also leave $W_C(x, y)$ invariant. Now

$$\left(\frac{1}{\sqrt{2}} \begin{pmatrix} -i & 1 \\ i & 1 \end{pmatrix}\right)^3 = \frac{1 + i}{\sqrt{2}} I$$

and so if $\tau := (1 + i)/\sqrt{2}$, then

$$W_C(x, y) = W_C(\tau x, \tau y) = \tau^n W_C(x, y).$$

But τ is an eighth root of 1, and we are forced to the conclusion that the length of the extended code is divisible by eight. □

This coding theory condition does not eliminate any plane that is not already excluded by the Bruck-Ryser-Chowla conditions. (It is possible to extend the coding theory arguments and hence rederive the conclusions of Bruck-Ryser-Chowla theorem.)

11.7 A 5-Design on 12 Points

We construct a 5-$(12, 6, 1)$ design. Let H be a Hadamard matrix of order 12. Since $H^T H = 12I$ we see that $\det(H) = 12^6 = 2^{12}3^6$, and so by Lemma 11.3.2 we have that $\mathrm{rk}_3(H) \geq 6$. On the other hand $HH^T = 0$

modulo three, and therefore the dimension of the row space of H over $GF(3)$ is at most six. Therefore $\mathrm{rk}_3(H) = 6$.

Let C the denote row space of H over $GF(3)$. This is a code of dimension six and length 12. Since $HH^T = 0$, it is self orthogonal and therefore it is a self-dual code.

We proceed in a number of steps.

If $z \in C$ then $\langle z, z\rangle$ must be zero; since the non-zero entries of z are ± 1 it follows that the weight of C is divisible by three.

Next we show that the minimum weight of a non-zero code word is six. Suppose $w \in C$ and $\mathrm{wt}(w) = 3$. Since $wH^T = 0$, there is a signed sum of three columns of H^T equal to zero. Since we can multiply rows of H by -1 without changing its row space, we may assume that the first of these columns is $\mathbf{1}$ and denote the others by x and y. As the weight distribution of $\mathrm{row}(H)$ does not change if we multiply columns of H by -1, we may suppose that $\mathbf{1} + x + y = 0$. But all entries of $x + y$ are even, we conclude that the minimum weight of C is at least six.

We claim that C contains at least 264 words of weight six. The point is that 264 is four times the number of pairs of rows. It is easy to see that for distinct rows x and y, all four vectors $\pm x \pm y$ have weight six. If

$$\pm x \pm y = \pm z \pm w$$

for four rows with $x \neq y$ and $z \neq w$, then the code formed by the column space of H contains a word of weight at most four. Therefore the pairs of distinct rows give rise to 264 words of weight six.

Finally we prove that the 132 supports of the words of weight six form a $5\text{-}(12, 6, 1)$ design. As such a design must have exactly 132 blocks, we need only show that no set of five is contained in two distinct supports. Assume by way of contradiction that x and y are words of weight six and S is a subset of size five such that

$$S \subseteq \mathrm{supp}(x) \cap \mathrm{supp}(y).$$

There are two cases. If $\mathrm{supp}(x) = \mathrm{supp}(y)$ then one of $x + y$ and $x - y$ has weight less then six. Othwerwise $S = \mathrm{supp}(x) \cap \mathrm{supp}(y)$. Then one of the code words $x + y$ and $x - y$ has weight at most five, again a contradiction. □

The code C is the *ternary Golay code* of length 12. It weight enumerator is

$$y^{12} + 264x^6y^6 + 440x^9y^3 + 24x^{12}.$$

11.8 Perfect Codes

The *ball of radius* e about a code word w is the set of words that are at distance at most e from w; we denote it by $B_e(w)$. The *packing radius* of a code C is the greatest integer e such that that the balls of radius e are pairwise disjoint. If C has packing radius e, then the minimum distance betwen two distinct code words is at least $2e+1$. The *covering radius* of C is the least integer r such that each word lies in $B_r(w)$, for some code word w. A code is *perfect* if its packing radius is equal to its covering radius.

If C is a code of length n over an alphabet of size q with packing radius e, then

$$|C| \leq \frac{q^n}{\sum_{i=0}^{e} \binom{n}{i}(q-1)^i};$$

this is the sphere packing bound, and follows trivially once we observe that the denominator is the size of a ball of radius e. A code is perfect if and only if equality holds in the sphere packing bound. We have the following interesting result, which we do not prove.

11.8.1 Lemma. *Suppose C is a perfect binary code of length n and packing radius e that contains the zero word. Then the supports of the code words of weight $2e+1$ form a design with parameters $(e+1)$-$(n, 2e+1, 1)$.* □

Note that in this lemma we do not require C to be a linear code. We offer one example. Let H be the matrix over $GF(2)$ with the distinct non-zero binary vectors of length k as its columns. Thus H is $k \times (2^k - 1)$. Let C be the kernel of H. You may show that the rows of H are linearly independent, whence

$$|C| = 2^{2^k - 1 - k}.$$

Since the columns of H are distinct and non-zero, there are no words in C with weight one or two and therefore the packing radius of C is at least 1. The ball of radius 1 about a word in a binary code of length n has size $n+1$, which is 2^k in our case. Hence the sphere-packing bound is tight and C is perfect. (In fact C is the binary Hamming code.) A perfect binary code with packing radius 1 gives rise to a Steiner triple system on $2^k - 1$ points. Examples are known that are not Hamming codes.

11.9 The Binary Golay Code

We construct a $[24, 12, 8]$-code over $GF(2)$. Let N be the incidence matrix of the symmetric 2-$(11, 6, 3)$ design (the complement of the Hadamard design). Define

$$\widehat{N} := \begin{pmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & N \end{pmatrix}.$$

and

$$G = \begin{pmatrix} I & \widehat{N} \end{pmatrix}.$$

11.9.1 Lemma. *The row space of G is a self dual binary code with with minimum distance eight.*

Proof. Since any two distinct blocks of the 2-$(11, 6, 3)$ design have exactly three points in common, it is easy to verify that any two rows of G are orthogonal. Since each row of G is doubly even, row(G) is a doubly even code. You may show that it does not contain any words of weight four. □

Since row(G) contains the vector $\mathbf{1}$ and since $\mathrm{wt}(\mathbf{1} + u) = n - \mathrm{wt}(u)$, the wieght enumerator $W_C(x, y)$ of our code has the form

$$y^{24} + Ax^8y^{16} + Bx^{12}y^{12} + Ax^{16}y^8 + x^{24}$$

where $A = A_8$ and $B = A_{12}$. As $W_C(1, 1) = |C|$ we have

$$B = 2^{12} - 2 - 2A.$$

We can now use MacWilliams' theorem to compute the number of words of weight two in the dual of C. Since this number is zero, we have an equation that determines A; in fact we find that $A = 759$.

11.9.2 Lemma. *The words of weight eight in a binary $[24, 12, 8]$-code form a 5-$(24, 8, 1)$ design.*

Proof. Let α be a subset of $V = \{1, \ldots, 24\}$ with size five. If x and y are two code words of weight eight with α in their support, then $x + y$ has weight at most six. Hence each 5-subset of V is contained in the support of at most one code word of weight eight, and so the number of such words is at most

$$\frac{\binom{24}{5}}{\binom{8}{5}} = 759$$

and, if equality holds, we have our 5-design.

Part II

Finite Geometry

Chapter 12

Projective Spaces

A *projective geometry* is an incidence structure such that:

(a) Any two distinct points lie on exactly one line.

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

(c) Every line contains at least three points.

The second condition is known as the *Veblen-Young axiom*. Clearly any projective plane is a projective geometry according to these axioms, but not much else is clear. You could also verify that the incidence structure with the 1-dimensional subspaces of a vector space V as points and 2-dimensional subspaces as lines does satisfy these axioms. We denote this structure by $\mathcal{P}(V)$.

12.1 Rank and Subspaces

A subset S of the points of a partial linear space is a *subspace* if any line that contains two points of S has all its points in S. Any line is a subspace. The intersection of any family of subspaces is a subspace and so, if P is a set of points, there is a unique minimal subspace that contains P. We call it the subspace *generated* by P. It follows that we can define the *join* $S_1 \vee S_2$ of two subspaces to be the subspace generated by $S_1 \cup S_2$.

If our incidence structure is $\mathcal{P}(V)$ for some vector space V, then a subspace in the sense just defined is a subspace of V in the sense used in linear

algebra. here the join is usually known as the sum, and denoted $S_1 + S_2$. You can show that $S_1 \vee S_2$ is the union of all lines in $\mathcal{P}(V)$ that contain points in S_1 and S_2. (It is immediate that the join contains all points on these lines, what requires proof is that there are no other points in the join.)

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

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

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

If, in addition

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

then we say the rank function is submodular.

A set equipped with a submodular rank function is called a *matroid*. A *flat* in a matroid is a subset F such that, if $p \notin F$ then $\mathrm{rk}(p \cup F) > \mathrm{rk}(F)$. A *combinatorial geometry* is a set P, together with a submodular rank function rk such that if $A \subseteq P$ and $|A| \leq 2$ then $\mathrm{rk}(A) = |A|$. Any combinatorial geometry can be regarded as a linear space with the flats of rank one as its points and the flats of rank two as its lines. The flats of rank three are its *planes*. The maximal proper flats of a combinatorial geometry are its *hyperplanes*; it is left as an exercise to show that these must all have the same rank. The rank of a combinatorial geometry is the maximum value of its rank function. We will always restrict ourselves to combinatorial geometries where the rank is finite, even if the point set is not.

The flats relative to a rank function form a lattice, such lattices are called *geometric lattices*. Note that in this case the join $F_1 \vee F_2$ of two flats is not (in general) equal to $F_1 \cup F_1$, although $F_1 \wedge F_1 = F_1 \cap F_2$. If for any two flats A and B we have

$$\mathrm{rk}(A \vee B) + \mathrm{rk}(A \wedge B) = \mathrm{rk}(A) + \mathrm{rk}(B)$$

we say that our rank function is *modular*. The lattices of subspaces of a vector space provides the most important example.

[*** rank for $AG(V)$ ***]

12.2 Projective Geometries and Planes

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

Proof. Let S be the set of all points which lie on a line joining p to a point of H. We will show that S is a subspace of $\mathcal{G}$. Suppose that ℓ is a line containing the points x and y from S. We have to show that each point of ℓ lies on a line joining p to a point of H.

By the definition of S, the point y is on line joining p to a point in H and if $x = p$ then this line must be ℓ.

If both x and y lie in H then $\ell \in H$, since H is a subspace.

Finally assume that x and y are both distinct from p and do not lie in H. It follows that both x and y lie on lines through p which meet H. Suppose that they meet H in x' and y' respectively. The line ℓ meets the line $p \vee x'$ and $p \vee y'$ in distinct points; therefore it must intersect $x' \vee y'$ in some point q. If u is a point on ℓ then the line $p \vee u$ meets $y \vee y'$ in p and $y \vee q$ in u. Hence it must meet the line $q \vee y'$, which lies in H. As u was chosen arbitrarily on ℓ, it follows that each point of ℓ lies on a line joining p to a point of H.

Thus we have shown that all points on ℓ lie in S, and so S is a subspace. Any subspace which contains both p and H must contain all points on the lines joining p to points of H. Thus S is the intersection of all subspaces containing p and H, i.e., $S = p \vee H$. □

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

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

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

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

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

12.3 Projective Geometries from Vector Spaces

We describe the most important construction of projective geometries. Let V a vector space with dimension at least two over a field $\mathbb{F}$. Let $\mathcal{P}(V)$ denote incidence structure formed by the 1-dimensional and 2-dimensional subspaces of V, where a 1-dimensional subspace is incident with the 2-dimensional subspaces that contain it.

We define a rank function on subsets of V by defining the rank of a subset to be the dimension of the subspace it spans. Here the flats are precisely the subspaces of V, and the join of two subspaces H and K is the subspace sum $H + K$. Since for two subspaces H and K we have

$$\dim(H + K) = \dim(H) + \dim(K) - \dim(H \cap K)$$

it is easy to verify that the rank function is submodular.

A projective space of rank two is a *projective line* while a space of rank three is a *projective plane*.

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

A such that $M = NA$. (The subspace will be determined uniquely by the reduced column-echelon form of M.)

Any invertible linear mapping of V to itself permutes the 1-dimensional and 2-dimensional subspaces of V and preserves incidence. Hence it gives rise to an automorphism of the projective space. Note though that the scalar matrices cI give rise to the identity automorphism on the projective space. We can also verify that any field automorphism gives rise to an automorphism of our projective space. The automorphism group of a projective space acts transitively on the set of subspaces of given rank.

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

The following lemma records a fundamental property of Desarguesian projective spaces.

12.3.1 Lemma. *Suppose V is a vector space of dimension at least three of the field $\mathbb{F}$. If ℓ is a line and H is a hyperplane in $\mathcal{P}(V)$, then $\ell \cap H \neq \emptyset$.*

Proof. Suppose the line is the subspace spanned by x and y and the hyperplane H is given by a vector a. If $a^T x = 0$ then x is on H, otherwise the vector

$$(a^T x)y - (a^T y)x$$

is on H (and on our line). □

You might also prove this using the formula for $\dim(H + K)$.

12.4 The Rank Function of a Projective Geometry

One of the most important properties of projective geometries is that their rank functions are modular. Proving this is the main goal of this section. Note that if p is a point and H a subspace in any linear space and $p \notin H$, then $\mathrm{rk}(p \vee H) \geq \mathrm{rk}(H) + 1$. We will use this fact repeatedly.

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

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

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

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

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

Proof. We will actually prove a more powerful result. Let H and K be two distinct maximal subspaces. Let h be point in $H \setminus K$ and let k be a point in $K \setminus H$. The line $h \vee k$ cannot contain a second point, h' say, of H since then we would have $k \in h \vee h' \subseteq H$. Similarly $h \vee k$ cannot contain a point of K other than k. By the third axiom for a projective geometry, $h \vee k$ must contain a point p distinct from h and k and, by what we have just shown, $p \notin H \cup K$. Since H and K are maximal $p \vee H = p \vee K$. By 12.2.2, each line through p must contain a point of H and a point of K. Using p we construct a mapping ϕ_p from H into K. If $h \in H$ then

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

If $\phi_p(h_1) = \phi_p(h_2)$ then the lines $p \vee h_1$ and $p \vee h_2$ have two points in common, and therefore coincide. This implies that they meet H in the same point and hence ϕ_p is injective.

If $k \in K$ then $k \vee p$ must contain a point h' say, of H. We have $\phi_p(h') = k$, whence ϕ_p is surjective. Thus we have shown that ϕ_p is a bijection.

We prove next that ϕ_p maps subspaces onto subspaces.

Let L be a subspace of H. Then $\phi_p(L)$ lies in $(p \vee L) \cap K$. Conversely, if $x \in (p \vee L) \cap K$ then x is on a line joining p to a point of L and so $x \in \phi_p(L)$. Hence $\phi_p(L) = (p \vee L) \cap K$. Since $p \vee L$ is a subspace, so is $(p \vee L) \cap K$. As ϕ_p is bijective on points, it must map distinct subspaces of H onto distinct subspaces of K. A similar argument to the above shows that ϕ_p^{-1} maps subspaces of K onto subspaces of H. Consequently we have shown that

ϕ_p induces an isomorphism from the lattice of subspaces of H onto the subspaces of K. This implies immediately that H and K have the same rank. (It is also worth noting that it implies that ϕ_p is a collineation—it must map subspaces of rank two to subspaces of rank two.) □

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

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

Proof. As $H \cap K$ is contained in both $p \vee K$ and H and as $p \in H$, it follows that $p \vee (H \cap K) \subseteq (p \vee K) \cap H$. Let x be a point in $(p \vee K) \cap H$. By 12.2.2, there is a point k in K such that $x \in p \vee k$. Now $p \vee k = p \vee x$ and so $k \in p \vee x$. Since $x \in H$ then this implies that $p \vee x \subseteq H$ and thus that k lies in H as well as K. Summing up, we have shown that if $x \in (p \vee K) \cap H$ then $x \in p \vee k$, where $k \in H \cap K$, i.e., that $x \in p \vee (H \cap K)$. □

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

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

Proof. We use induction on $\mathrm{rk}(H) - \mathrm{rk}(H \cap K)$. Suppose first that this difference is equal to one. This implies $H \cap K$ is maximal in H and accordingly

$$\mathrm{rk}(H) - \mathrm{rk}(H \cap K) = 1. \tag{12.4.1}$$

If $p \in H \setminus K$ then, using the maximality of $H \cap K$ in H, we find that $p \vee (H \cap K) = H$ and $H \vee K = p \vee K$. By 12.4.2 it follows that K is maximal in $H \vee K$ and so

$$\mathrm{rk}(H \vee K) - \mathrm{rk}(K) = 1. \tag{12.4.2}$$

Subtracting (12.4.1) from (12.4.2) and rearranging yields the conclusion of the Theorem.

Assume now that $H \cap K$ is not maximal in H. Then we can find a point $p \in H \cap K$ such that $p \vee (H \cap K) \neq H$. Suppose $L = p \vee (H \cap K)$. Then $H \cap K$ is maximal in L (by 12.4.2) and so, by what we have already proved,

$$\mathrm{rk}(L \vee K) + \mathrm{rk}(L \cap K) = \mathrm{rk}(L) + \mathrm{rk}(K). \tag{12.4.3}$$

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

$$\mathrm{rk}(H \vee (L \vee K)) + \mathrm{rk}(H \cap (L \vee K)) = \mathrm{rk}(L \vee K) + \mathrm{rk}(H). \qquad (12.4.4)$$

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

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

Furthermore $H \vee (L \vee K) = H \vee K$, and so (12.4.4) can be rewritten as

$$\mathrm{rk}(H \vee K) + \mathrm{rk}(L) = \mathrm{rk}(L \vee K) + \mathrm{rk}(H). \qquad (12.4.5)$$

Since $L \cap K = H \cap K$, we can now derive the theorem by adding (12.4.3) to (12.4.5) and rearranging. □

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

12.5 Duality

Let H and K be two maximal subspaces of a projective geometry with rank n. Then $\mathrm{rk}(H \vee K) = n$ and from 12.4.5 we have

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

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

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

Proof. We first show that each line in the dual lies on at least three points. Let K be space of rank $n-2$ and let H_1 be a hyperplane which contains it. Since H_1 is not the whole space, there must be point p not in it. Then

K is maximal in $p \vee K$ and so $p \vee K$ is a subspace of rank $n-1$ on k. It is not equal to H, because p is in it. Now choose a point q in $H \setminus K$. The line $p \vee q$ must contain a third point, x say. If $x \in H$ then $p \in x \vee q \subseteq H$, a contradiction. Similarly x cannot lie in K and so it follows that $x \vee K$ is a third subspace of rank $n-1$ on K.

Now we should verify the second axiom. However we will show that any two subspaces of rank $n-2$ intersecting in a subspace of rank $n-3$ lie in a subspace of rank $n-1$. This implies that any two lines in the dual which line a subspace of rank three must intersect, and so all rank three subspaces are projective planes. An appeal to 12.2.3 now completes the proof. So, suppose that K_1 and K_2 are subspaces with rank $n-2$ which meet in a subspace of rank $n-3$. Then

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

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

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

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

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

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

12.6 Some Counting

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

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

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

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

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

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

12.6.1 Theorem. *Let V be a vector space of dimension n over a field of finite order q. Then the number of subspaces of V with dimension k is $\begin{bmatrix} n \\ k \end{bmatrix}$.*

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

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

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

Hence

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

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

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

This yields the theorem. □

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

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

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

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

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

We have $H + W = V$ and therefore

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

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

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

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

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

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

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

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

12.6.2 Lemma. *Let V be a vector space of dimension n over a field of order q, and let U be a subspace of dimension k. The number of subspaces of V with dimension ℓ that are skew to U is $q^{k\ell}\begin{bmatrix} n-k \\ \ell \end{bmatrix}$.*

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

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

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

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

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

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

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

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

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

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

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

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

Exercises

1. Determine the projective spaces that are not thick.

2. If H, K and L are subspaces of a projective geometry and $L \leq H$, show that
$$H \cap (L \vee K) = L \vee (H \cap K)$$
(This is equivalent to requiring that $H \cap (L \vee K) = (H \cap L) \vee (H \cap K)$, and is known in lattice theory as the *modular law*.)

3. Let $\mathcal{A}$ be an affine space of rank d and order two. Show that each parallel class of lines determines a permutation of the points of $\mathcal{A}$ with all orbits of length two. Show that this set of permutations, together with the identity, forms an abelian group of order 2^d where each non-identity element has order two. (Such a group is said to be *elementary abelian* of exponent two.)

4. Prove that each elementary abelian group of exponent two determines an affine geometry.

5. Prove that in a projective space of rank r, any subspace of rank at most $r-2$ is the intersection of the subspaces that cover it.

6. Prove that the number of invertible $d \times d$ matrices over $GF(q)$ is equal to $q^{\binom{d}{2}}(q-1)^d[d]!$.

7. Let V be a vector space of dimension n over $GF(q)$ and let U be a subspace of dimension k. Determines the number of subspaces of W with dimension ℓ such that $U \cap W = 0$.

8. Let $\mathcal{P}$ be the projective space of rank $d-1$ over $GF(q)$. Let S be a set of points that meets every subspace of rank e. Show that $|S|$ contains at least as many points as a subspace of rank $d-e+1$; if equality holds prove that S is a subspace.

9. Prove that the number of invertible $d \times d$ matrices over $GF(q)$ is equal to $q^{\binom{d}{2}}(q-1)^d[d]!$.

Chapter 13

Affine Spaces

13.1 Affine Geometries

Let $\mathcal{G}$ be a projective geometry and let H be a hyperplane in it. If S is a set of points in $\mathcal{G} \setminus H$, define $\mathrm{rk}_H(S)$ to be $\mathrm{rk}(S)$. This can be shown to be a submodular rank function on the points not on H, and the combinatorial geometry which results is an *affine geometry*. (It will sometimes be denoted by $\mathcal{G}^H$.) The flats of $\mathcal{A}$ are the subsets of the form $K \setminus H$, where K is a flat/subspace of $\mathcal{G}$. They will be referred to as *affine subspaces*; these are all linear subspaces. However, in some cases there will be linear subspaces which are not flats. (This point will be considered in more detail at the start of the following section.)

If K_1 and K_2 are two subspaces of $\mathcal{G}$ such that $K_1 \cap K_2 \subseteq H$ and

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

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

extended uniquely to a parallel class. For given two parallel lines, we can identify the point p on H where the meet; the remaining lines in the parallel class are those that also meet H at p.

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

We consider an example to illustrate some issues. Let V be a vector space of dimension d over $GF(2)$. We can form an incidence structure with the vectors in V as points and the cosets of the 1-dimensional subspaces as lines. Then each pair of points is a line and so each subset of points is a subspace. We could view this as a combinatorial geometry where each subset is a flat and the rank of a subset S is $|S|$; this means we that we are ignoring a lot of structure. However there is a second rank function available: define the rank of a set of vectors to be the dimension of the affine space that they span, plus one. Now the flats are the affine subspaces. It is this rank function that we will use.

13.2 Axioms for Affine Spaces

There is a difficulty in providing a set of axioms for affine spaces, highlighted by the following. Consider the projective plane $PG(2,2)$. Removing a line from it gives the affine plane $PG(2,2)$ which has four points and six lines; each line has exactly two points on it. (Thus we could identify its points and

lines with the vertices and edges of the complete graph K_4 on four vertices.) This is a linear space but, unfortunately for us, it has rank four: any set of three points is a subspace of rank three. More generally, any subset of the points of $AG(n, 2)$ is a subspace of $AG(n, 2)$, viewed as a linear space. However not all subspaces are flats.

One set of axioms for affine spaces has been provided by H. Lenz. An *affine space* is an incidence structure equipped with an equivalence relation on its lines, called parallelism and denoted by $\|$, such that the following hold.

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

(b) Given any line ℓ and point p not on ℓ, there is a unique line ℓ' through p and parallel to ℓ.

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

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

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

(f) Any line has at least two points.

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

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

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

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

13.3 Affine Spaces in Projective Space

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

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

Proof. Let W denote the union described. Since the subspace $D \vee \pi$ is the join of D and any line in π which does not meet D, no line in π which does not meet D can be coplanar with it. Hence no line in π is parallel to D.

If x is point in W which is not on D then $x \vee D$ is a plane. Since $x \in W$, there is a plane containing x and D which meets π in a line. Thus $x \vee D$ must meet π in line, l say. As x is not on l, there is unique line, l' say, parallel to it through x. Since D is not parallel to l, it is not parallel to l'. Therefore D meets l'. We will denote the point of intersection of D with l' by $d(x)$. Now suppose that x and y are distinct points of W. We seek to show that any point on $x \vee y$ lies in W. There are unique lines through x and y parallel to D; since they lie in $x \vee D$ and $y \vee D$ respectively they meet π in points x' and y'. If u is a point on $x \vee y$ then the unique line through u parallel to D must intersect $x' \vee y'$. Hence u lies in the plane spanned by this point of intersection and D, and so $u \in W$. □

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

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

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

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

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

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

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

Let $\mathcal{A}$ be an affine geometry. We show how to embed it in a projective geometry. Assume that the rank of $\mathcal{A}$ is at least three. (If the rank is less than three, there is almost nothing to prove.) Let P be a set with cardinality equal to the number of parallel classes. We begin by adjoining P to the point set of $\mathcal{A}$. If a line of $\mathcal{A}$ lies in the i-th parallel class, we extend it by adding the i-th point of P. It is straightforward to show that each plane in $\mathcal{A}$ has now been extended to a projective plane. Each plane in $\mathcal{A}$ determines a set of parallel classes, and thus a subset of P. These subsets are defined to be lines of the extended geometry; the original lines will be referred to as *affine* lines if necessary. Two points a and b of $\mathcal{A}$ are collinear with a point p of P if and only the line $a \vee b$ is in the parallel class associated with p. With the additional points and lines as given, we now have a new incidence structure $\mathcal{P}$. We must verify that it is a linear space.

Let a and b be two points. If these both lie in $\mathcal{A}$ then there is a unique line through them. If $a \in \mathcal{A}$ and $b \in P$ then there is a unique line in the parallel class determined by b which passes through a. Finally, suppose that a and b are both in P. Let l be a line in the parallel class determined by a. If x is an affine point in l then there is unique line in the parallel class of b passing through it. With l, this line determines a plane which contains all

the lines in b which meet l. This shows that each line l in a determines a unique plane.

We claim that it is a projective space. This can be proved by showing that each plane in $\mathcal{P}$ is projective. The only difficult case is to verify that the planes contained in P are projective. Each plane of P corresponds to a subspace of $\mathcal{A}$ with rank four, so studying the planes of P is really studying these subspaces of $\mathcal{A}$. The planes contained in P are projective planes if every pair of lines in them intersect. Thus we must prove that if σ and π are two planes of $\mathcal{A}$ contained in a subspace of rank four, then there is line in σ parallel to π. There are two cases to consider. Suppose first that $\sigma \cap \pi = \emptyset$. Then, by 13.3.2, any line in σ is parallel to a line in π, and therefore there is a point in P lying on both the lines determined by σ and π. Suppose next that σ and π have a point in common. Then, by 13.3.3, these planes must have a line in common and so the parallel class containing it lies on the lines in P determined by them. This completes the proof that all thick affine spaces are projective spaces with a hyperplane removed.

It is not at all difficult to show that an affine geometry over $GF(2)$ is isomorphic to the affine space associated the vector space of the same dimension over $GF(2)$. Consequently any affine geometries of order two has a unique embedding into a projective space.

In the next chapter we will use our axiomatic characterization of affine spaces to show that all projective spaces of rank at least four have the form $\mathbb{P}(n, \mathbb{F})$, that is, are projective spaces over some skew field.

13.4 Characterizing Affine Spaces by Planes

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

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

Proof. We verify that the axioms for thick affine spaces hold. Since the second of these axioms is a condition on planes, it is automatically satisfied.

Thus we need only prove that parallelism is an equivalence relation on the lines of $\mathcal{A}$. If π is a plane and D a line meeting π in the point a, we define $W = W(\pi, D)$ to be the union of the point sets of the planes which contain D and meet π in a line.

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

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

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

Suppose that l_1, l_2 and l_3 are lines in $\mathcal{A}$, with $l_1 \parallel l_2$ and $l_2 \parallel l_3$. Let π be the plane $l_1 \vee l_2$, let D be a line joining a point b on l_3 to a point a in l_2 and let $W = W(\pi, D)$. Since $b \in W$ and W is a subspace, the plane $b \vee l_1$ lies in

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

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

13.5 Affine Spaces

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

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

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

$$v = \sum_i a_i v_i$$

where

$$\sum_i a_i = 1.$$

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

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

If u and v are distinct vectors, then the set

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

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

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

13.6 Coordinates

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

Now suppose $\mathcal{P}$ is the projective space associated to $\mathbb{F}^n$. Two non-zero vectors x and y represent the same point if and only if there is a non-zero scalar a such that $y = ax$. Thus a point is an equivalence class of non-zero vectors. As usual it is often convenient to represent an equivalence class by one of its elements. Here there is no canonical choice, but we could take the representative to be the vector with first non-zero coordinate equal to 1. Normally we will **not** do this.

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

Defining functions on projective space is trickier, because each point is represented by a set of vectors. However if p is a homogeneous polynomial

in n variables with degree k and $x \in \mathbb{F}^n$, then

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

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

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

Chapter 14

Collineations and Perspectivities

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

14.1 Collineations of Projective Spaces

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

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

Thus τ does not induce a linear mapping of V onto itself, but it does map subspaces to subspaces, and therefore does induce a collineation. If we apply any sequence of linear mappings and field automorphisms to $\mathbb{P}(n, \mathbb{F})$

then we can always obtain the same effect by applying a single linear mapping followed by a field automorphism (or a field automorphism then a linear mapping). The composition of a linear mapping and a field automorphism is called a *semi-linear* mapping. The set of all collineations obtained by composing linear mappings and field automorphisms is called the group of projective semi-linear transformations of $\mathbb{P}(n, \mathbb{F})$, and is denoted by $P\Gamma L(n+1, \mathbb{F})$. It contains $PGL(n, \mathbb{F})$ as a normal subgroup of index equal to $|\mathrm{Aut}(\mathbb{F})|$. (If $\mathbb{F}$ is finite of order p^m, where p is prime, then $\mathrm{Aut}(\mathbb{F})$ is a cyclic group of order m generated by the mapping which sends an element x of $\mathbb{F}$ to x^p.) We can now state the "fundamental theorem of projective geometry".

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

Proof. See the end of **??**. □

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

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

Proof. Exercise. □

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

To compute the order of $PGL(n, \mathbb{F})$ when $\mathbb{F}$ is finite with order q, we first compute the order of $GL(n, \mathbb{F})$. This is just the number of non-singular $n \times n$ matrices over $\mathbb{F}$. We can construct such matrices one row at a time. The number of possible first rows is $q^n - 1$ and, in general, the number of

possible $(k+1)$-th rows is the number of vectors not in the span of the first k rows, that is, it is $q^n - q^k$. Hence

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

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

A *k-arc* in a projective geometry of rank n is a set of k points, no n of which lie in a hyperplane. To construct an $(n+1)$-arc in $\mathbb{P}(n-1,\mathbb{F})$, take a basis $x_1,\ldots,x_n$ of $V(n,\mathbb{F})$, together with a vector y of the form $\sum_i a_i x_i$, where none of the a_i are zero. The linear transformation which sends each vector x_i to $a_i x_i$ maps $\sum x_i$ to $\sum a_i x_i$. Hence the subgroup of $PGL(n,\mathbb{F})$ fixing each of $x_1,\ldots,x_n$ acts transitively on the set of points of the form $\sum a_i x_i$, where the a_i are non-zero. It is also possible to show that a collineation of $\mathbb{P}(n-1,\mathbb{F})$ which fixes each point in an $(n+1)$-arc is the identity. (The proof of this is left as an exercise.) Together these statements imply that the subgroup of $GL(n,\mathbb{F})$ fixing each of $x_1,\ldots,x_n$ acts regularly on the set of points of the form $\sum a_i x_i$, where the a_i are non-zero, and hence that it has order $(q-1)^n$. The subgroup of $P\Gamma L(n+1,\mathbb{F})$ fixing each point in an $(n+1)$-arc can be shown to be isomorphic to the automorphism group of the field $\mathbb{F}$. (See Hughes and Piper [].)

14.2 Perspectivities and Projections

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

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

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

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

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

If $\mathcal{G}$ has rank n then the hyperplanes H and K meet in a subspace of rank $n-2$ and each point in this subspace is fixed by $\phi_p\psi_q$. All lines through the point $(p \vee q) \cap H$ are also left fixed by $\phi_p\psi_q$. As $H \cap K$ is a hyperplane in H, it follows that $\phi_p\psi_q$ is a perspectivity. We will make considerable use of these perspectivities in proving that all projective geometries of rank at least four arise as the 1- and 2-dimensional subspaces of a vector space.

We provide a class of linear mappings of a vector space which induce perspectivities of the associated projective space $\mathbb{P}(n, \mathbb{F})$. Let $V = V(n, \mathbb{F})$ and let H be a hyperplane in V. A linear mapping τ of V is a *transvection* with *axis* H if $x\tau = x$ for all x in H, and $x\tau - x \in H$ for all x not in H. They are easy to construct. First, choose a bilinear form $\langle \cdot, \cdot \rangle$ on V. Next choose non-zero vectors h and a such that $\langle h, a \rangle = 0$ and define $\tau_{h,a}$ by setting

$$x\tau_{h,a} = x - \langle x, h \rangle a.$$

Then x is fixed by $\tau_{h,a}$ if and only if $\langle h, x \rangle = 0$. Thus $\tau_{h,a}$ fixes all points of the hyperplane with equation $\langle h, x \rangle = 0$. If x is not fixed by $\tau_{h,a}$ then $x\tau_{h,a} - x$ is a multiple of a and, since $\langle h, a \rangle = 0$, it follows that a lies in the hyperplane of points fixed by $\tau_{h,a}$. As $\tau_{h,a}$ fixes a, it follows that it also fixes all the 2-dimensional subspaces $a \vee x$.

14.3 Groups of Perspectivities

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

Proof. Assume first that τ fixes some point c not in H and let l be a line through c. Then l must meet H in some point, x say. As $x \in H$, it is fixed

by τ and thus τ fixes two distinct points of l. This implies that l is fixed by τ.

Assume now that there are no points off H fixed by τ. Let p be a point not in H and let $l = p \vee p\tau$. Once again l must intersect H in some point, x say. As τ fixes x and maps p in l to $p\tau$ in l, it follows that it fixes l.

Let q be a point not on H or l. The plane $\pi = q \vee l$ meets H in a line l' (why?). Since τ fixes the distinct lines l and l' from π, it also fixes π. This implies that $q\tau \in \pi$. Now $q\tau \neq q$, since $q \notin H$, and so the $q \vee q\tau$ is a line in π. Hence it intersects l' and, since $l' \subseteq H$, the point of intersection is fixed by τ. Therefore $q \vee q\tau$ is fixed by τ. The line $q \vee q\tau$ must intersect l in some point, c say. As $q \vee q\tau$ and l are both fixed by τ, so is c. Therefore $c \in H$ and so $c = H \cap l = x$. Thus we have shown that the lines $q \vee q\tau$, where $q \notin H$, all pass through the point c in H. From this it follows that all lines through c are fixed by τ. □

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

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

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

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

Proof. Only the last claim needs proof. Suppose a and b are distinct points off H fixed by τ. If p is a third point, not in H, then τ fixes the point

$H \cap pa$ as well as a. hence τ fixes pa and similarly it fixes pb. Therefore $p = pa \cap pb$ is fixed by τ. This shows that τ fixes all points not in H. □

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

Proof. The challenge is to show that the centre of $\tau_1\tau_2$ is on the line joining the centers of τ_1 and τ_2.

Denote the respective centres of τ_1 and τ_2 by c_1 and c_2. If $c_1 = c_2$, then c_1 is the center of $\tau_1\tau_2$ and we are done. So we may assume throughout that $c_1 \neq c_2$ and that neither τ_1 nor τ_2 is the identity.

We note that any line that is fixed by a perspectivity and which is not contained in its axis must contain the center of the perspectivity.

Suppose first the τ_1 is a homology, i.e., $c_1 \notin H$. Then $c_1\tau_2 \neq_1$ and the line $m = c_1 \vee c_1\tau_2$ is fixed by τ_2 (because its intersection with H is fixed by τ_2). It follows that m contains the centre of τ_2. Since m contains the center of τ_1, it is also fixed by τ_1. We conclude that $\tau_1\tau_2$ fixes m and therefore the center of $\tau_1\tau_2$ is on m. As $m = c_1 \vee c_2$, we are done.

So we may assume that τ_1 and τ_2 are elations. If p is a point off H fixed by $\tau_1\tau_2$, then

$$p\tau_1 = p\tau_2^{-1}$$

and therefore the line $p \vee p\tau_1 = p \vee p\tau_2^{-1}$ is fixed by τ_1 and τ_2. Hence this line must contain the centers of both elations and therefore $c_1 = c_2$. We conclude that $\tau_1\tau_2$ does not fix any point of H and therefore it is an elation.

If our projective space has rank three, then $H = c_1 \vee c_2$ and so the center of $\tau_1\tau_2$ is on $c_1 \vee c_2$ as required. Suppose the rank is greater then three and let q be a point off H. Since τ_1 fixes $q \vee c_1$, it fixes the plane $\pi = q \vee (c_1 \vee c_2)$. Since τ_2 fixes $q \vee c_2$ it also fixes π. Hence $\tau_1\tau_2$ fixes π; as it induces a perspectivity of π and the center p of this must be on $\pi \cap H$., it must fix all lines in π through some point on $c_1 \vee c_2$. We conclude that p is the center of $\tau_1\tau_2$. □

One consequence of the previous lemma is that the product of two elations with axis H is always a perspectivity with axis H and centre on H, that is, it is an elation. It is possible for the product of two homologies with axis H to be an elation—with its centre the point of intersection of H with the line through the centres of the homologies.

14.4 Transitivity Conditions

We now come to an important definition. Let p be a point and H a hyperplane in the projective geometry $\mathcal{G}$ and let Γ be a group of collineations of $\mathcal{G}$. Let $\Gamma(p, H)$ denote the subgroup of Γ formed by the perspectivities with centre p and axis H. We say that Γ is (p, H)-*transitive* if, for any line ℓ through p which is not contained in H, the subgroup $\Gamma(p, H)$ acts transitively on the set of points of ℓ which are not on H and not equal to p. If $\mathrm{Aut}(\mathcal{G})$ is itself (p, H)-transitive then we say that $\mathcal{G}$ is (p, H)-*transitive*.

This is a reasonable point to explain some group theoretic terms as well. If Γ is a permutation group acting on a set S and $x \in S$ then Γ_x is the subgroup of Γ formed by the permutations which fix x. Recall that the length of the orbit of x under the action of Γ is equal to the index of Γ_x in Γ. The group Γ is transitive if it has just one orbit on S. It acts *fixed-point freely* on S if the only element which fixes a point of S is the identity, that is, if Γ_x is the trivial subgroup for each element x in S. In this case each orbit of Γ on S will have length equal to $|\Gamma|$ (and so $|\Gamma|$ divides $|S|$ when everything is finite).

Suppose that Γ is the group of all perspectivities of $\mathcal{G}$ with centre p and axis H. Let q be a point not in H and distinct from p. If an element γ of Γ fixes q then it is the identity. For since $q \notin H$ and since γ fixes each point in H it fixes all lines joining q to a point in H. But as H is a hyperplane, this means that it fixes all lines through q. Hence q must be the centre of γ, and so $q = p$. This contradiction shows that Γ must act fixed-point freely on the points of $\mathcal{G}$ not in $H \cup p$. In particular, for any line l, we see that Γ acts fixed-point freely on the points of $l \setminus p$ not in H. (Since $p \in l$, the line l must be fixed as a set by Γ.) Therefore if $\mathcal{G}$ is finite and $p \notin H$ then $|\Gamma|$ must divide $|l| - 2$, and if $p \in H$ then $|\Gamma|$ divides $|l| - 1$.

14.5 Desarguesian Projective Planes

Let $\mathcal{P}$ be a projective plane, with p a point and ℓ a line in it. The condition that $\mathcal{P}$ be (p, ℓ)-transitive can be expressed in a geometric form. A *triangle* in a projective plane is a set of three non-collinear points $\{a_1, a_2, a_3\}$, together with the lines $a \vee b$, $b \vee c$ and $c \vee a$. These lines are also known as the *sides* of the triangle. For convenience we will now begin to abbreviate expressions such as $a \vee b$ to ab. Two triangles $\{a_1, a_2, a_3\}$ and $\{b_1, b_2, b_3\}$

are said to be *in perspective from a point* p if the three lines a_1b_1, a_2b_2 and a_3b_3 all pass through p. They are in *perspective from a line* ℓ if the points $a_1a_2 \cap b_1b_2$, $a_2a_3 \cap b_2b_3$ and $a_3a_1 \cap b_3b_1$ all lie on ℓ. We have the following classical result, known as *Desargues' theorem*.

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

Proof. Wait. □

A projective plane is (p, ℓ)-*Desarguesian* if:

- whenever two triangles $\{a_1, a_2, a_3\}$ and $\{b_1, b_2, b_3\}$ are in perspective from p, and
- both $a_1a_2 \cap b_1b_2$ and $a_2a_3 \cap b_2b_3$ lie on ℓ,

then $a_3a_1 \cap b_3b_1$ lies on ℓ. We call a plane *Desarguesian* if it is (p, ℓ)-Desarguesian for all points p and lines ℓ. Since the projective planes over fields are all Desarguesian, by the previous theorem, this concept is quite natural. However we will see that a plane is Desarguesian if and only if it is of the form $\mathbb{P}(2, \mathbb{F})$ for some skew-field $\mathbb{F}$.

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

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

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

need to construct a perspectivity with centre p and axis ℓ which sends x to y. If a is a point not on px or ℓ, define

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

and if $a \in \ell$, set $a\tau$ equal to a. As thus defined, τ is a permutation of the point set of the affine plane obtained by deleting px from $\mathcal{P}$.

We will prove that it is a collineation of this affine plane, and hence determines a collineation of $\mathcal{P}$ fixing px. Since $a\tau \in pa$, the mapping τ fixes the lines through p. Hence, if τ is a collineation then it is a perspectivity with centre and axis in the right place. Suppose that a and b are two distinct points of $\mathcal{P}$ not on px. If $b \in xa$ then $ax = ab$ and $((ax \cap \ell) \vee y) = ((ab \cap \ell) \vee y)$, implying that $b\tau$ is collinear with $y = x\tau$ and $a\tau$. Conversely, if $b\tau$ is collinear with y and $a\tau$ then b must be collinear with x and a.

Thus we may assume that x, a and b are not collinear. Then $\{x, a, b\}$ and $\{y, a\tau, b\tau\}$ are two triangles in perspective from the point p. By construction xa meets $y \vee a\tau$ and xb meets $y \vee b\tau$ on ℓ. Therefore $a \vee b$ must meet $a\tau \vee b\tau$ on ℓ. Let u be a point on ab. Then, applying Desargues' theorem a second time, we deduce that au and $a\tau \vee u\tau$ meet on ℓ. Since $au = ab$, they must actually meet at $ab \cap \ell$. Therefore

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

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

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

Proof. Let π be a plane in $\mathcal{G}$ and let p a point and ℓ a line in π. Let a and b be distinct points on a line in π through p, neither equal to p or on ℓ. Let σ be a second plane meeting π in ℓ and let v be a point not in $\pi \vee \sigma$ but not in π or σ. If $x \in \pi$ then $v \vee x$ must meet σ in a point. The mapping sending x to $(v \vee x) \cap \sigma$ is collineation ϕ_v from π to σ. The line ba' is contained in the plane $p \vee a \vee v$, as is pv. Hence ba' meets pv in a point, w say. Note that w cannot lie in π or σ. Hence it determines a collineation ϕ_w from σ to π which maps a' to b. Both ϕ_v and ϕ_w fix each point in ℓ, and so their composition is a collineation of π which fixes each point of ℓ and maps a to b. This shows that π is a (p, ℓ)-transitive plane. As our choice of p and ℓ was arbitrary, it follows from the previous two results that all planes in $\mathcal{G}$ are Desarguesian.

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

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

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

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

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

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

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

14.6 Translation Groups

Let H be a hyperplane in the projective geometry $\mathcal{G}$. (We assume that $\mathcal{G}$ has rank at least three.) The ordered pair $(\mathcal{G}, H)$ is an affine geometry and an elation of $\mathcal{G}$ with axis H and centre on H is called a *translation*. From Lemma 3.1, it follows that the set of all translations form a group. We are going to investigate the relation between the structure of $\mathcal{A}$ and this group. Some group theory must be introduced. A group Γ is *elementary abelian* if it is abelian and its non-identity elements all have the same order. If Γ is elementary abelian then so is any subgroup. As any element generates

a cyclic group, and as the only elementary abelian cyclic groups are the groups of prime order, all non-zero elements of a finite elementary abelian group must have order p, for some prime p. The group itself thus has order p^n for some n. We will usually use multiplication to represent the group operation, and consequently refer to the 'identity element' rather than the 'zero element'. (There will be one important exception, when we consider endomorphisms.) If H and K are subsets of the group Γ then we define

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

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

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

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

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

This shows that $\beta^{-1}\alpha\beta \in T(p)$. In other words, $T(p)$ is normalized by the elements of $T(q)$. If $\beta^{-1}\alpha\beta \in T(p)$ then the commutator $\alpha^{-1}\beta^{-1}\alpha\beta$ must also lie in $T(p)$. A similar argument shows that $\alpha^{-1}\beta^{-1}\alpha \in T(q)$. Accordingly $\alpha^{-1}\beta^{-1}\alpha\beta$ also lies in $T(q)$. As $T(p) \cap T(q) = \langle 1 \rangle$, it follows that $\alpha^{-1}\beta^{-1}\alpha = 1$. Consequently $\alpha\beta = \beta\alpha$. (In other words, two non-identity elations with the same axis and distinct centres commute.)

We now show that $T(p)$ is abelian. Let α' be a second non-identity element of $T(p)$. Then $\alpha'\beta$ is an elation. If its centre is p then β must belong to $T(p)$. Thus its centre is not p. Arguing as before, but with $\alpha'\beta$ in place of β, we deduce that α and $\alpha'\beta$ commute. This implies in turn that α and α' commute. Finally, assume that α is an element of $T(p)$ with order

m. If $\beta^m \neq 1$ then

$$(\alpha\beta)^m = \alpha^m\beta^m = \beta^m \in T(q). \tag{14.6.1}$$

Since $\alpha\beta$ is an elation with axis H, so is $(\alpha\beta)^m$, and (14.6.1) shows that its centre is q. Therefore the centre of $\alpha\beta$ is q and so $\alpha\beta \in T(q)$. Since $\beta \in T(q)$, we infer that α also lies in $T(q)$. This is impossible, and forces us to conclude that $\beta^p = 1$. Thus we have proved that two non-identity elations with distinct centres must have the same order. It is now trivial to show that $T(H)$ is elementary abelian. □

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

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

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

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

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

14.7 Geometric Partitions

Assume now that $\mathcal{G}$ is a projective geometry which is (p, H)-transitive for all points p on the hyperplane H, e.g., any projective geometry with rank at least four, or any Desarguesian plane. Then $T(H)$ is an elementary abelian group and the subgroups $T(p)$, where $p \in H$, partition its non-identity elements. In fact $T(H)$, together with the subgroups $T(p)$, completely determines $\mathcal{G}$. The connection is quite simple: the elements of $T = T(H)$

correspond to the points of $\mathcal{G} \setminus H$ and the cosets of the subgroups $T(p)$ are the lines. The correspondence between points and elements of T arise as follows. Let o be a point not in H. We associate with the identity of T. If a is a second point not on H then there is a unique elation τ_a with axis H and centre $H \cap oa$ which maps o to a. Then the map $a \mapsto \tau_a$ is a bijection from T to the points of $\mathcal{G}^H$. If l is a line of $\mathcal{G}^H$ then then the affine points of l are an orbit of $T(l \cap H)$, and conversely, each such orbit is a line.

This leads us naturally to conjecture that an elementary abelian group T, together with a collection of subgroups T_i $(i = 1, \ldots, m)$ such that the sets $T_i \setminus 1$ partition $T \setminus 1$, determines an affine geometry. This conjecture is wrong, but easily fixed. Let T be an elementary abelian group. A collection of subgroups T_i $(i = 1, \ldots, m)$ is a *geometric partition* of T if

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

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

A set of subgroups for which (a) holds is called a *partition* of T, although it is not quite. The partitions we have been studying are all geometric. For $T(p)T(q) = T(p \vee q)$ and so if $\tau \in T(r) \cap T(p)T(q)$ then r must lie on $p \vee q$ and so $T(r) \leqslant T(p)T(q)$. A geometric partition of an elementary abelian group determines an affine geometry $\mathcal{G}^H$. We take the affine points to be the elements of T and the lines to be the cosets of the subgroups T_i. This gives us a linear space. Showing that this is an affine geometry is left as an exercise.

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

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

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

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

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

$$T_i+((T_i+T_k)\cap T_j)=(T_i+T_k)\cap(T_i+T_j). \tag{14.7.1}$$

To prove this, note first that both terms on the left hand side are contained in the right hand side. Conversely, if u belongs to the right hand side then we can write it both as $x+y$ where $x\in T_i$ and $y\in T_j$, and as $x'+z$ where $x'\in T_i$ and $z\in T_k$. Since $x+y=x'+z$ we have $y=-x+x'+z\in(T_i+T_k)$ and so $y\in(T_i+T_k)\cap T_j$. If $T_k\le T_i+T_j$ then the right hand side of (14.7.1) is equal to T_i+T_k while, since the partition is geometric, the left hand side equals T_i or T_i+T_j. As $T_k\neq T_i$, this proves the lemma. □

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

Proof. Let K be the set of endomorphisms referred to. We show first that the elements of K are injective. Suppose that $\alpha\in K$ and $x\alpha=0$ for some element x of T. Assume that x is a non-zero element of T_1 and let y be a non-zero element of T_i for some i. Then

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

and therefore $(x+y)\alpha$ must lie in T_i, since $y\alpha$ does. On the other hand, $x+y$ cannot lie in T_i, and therefore $(x+y)\alpha=0$. This shows that $y\alpha=0$. As our choice of y in T_i was arbitrary, it follows that each element of T_i is mapped to zero and, as our choice of i was arbitrary, that $(T\setminus T_i)\alpha=0$.

Since $y\alpha = 0$, we may also reverse the role of x and y in the first step of our argument and hence deduce that $T_1\alpha = 0$. Thus we have proved that if α is not injective then it is the zero endomorphism.

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

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

14.8 The Climax

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

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

Proof. Let $T = T(H)$ and let K be the skew field of endomorphisms of the geometric partition determined by the subgroups $T(p)$, where $p \in H$. The non-zero elements of K form a group isomorphic to the group of all homologies of $\mathcal{G}$ with axis H and centre some point o off H. Since K is a skew field, we can view T as a vector space (over K) and the subgroups $T(p)$ as subspaces. As T acts transitively on the points of $\mathcal{G}$ not in H, it follows that $\mathcal{G}$ is (o, H)-transitive. This implies that $K \setminus 0$ acts transitively

on the non-identity elements of $T(p)$, and hence that $T(p)$ is 1-dimensional subspace of T. Consequently the affine geometry $\mathcal{G}^H$ has as its points the elements of the vector space T, and as lines the cosets of the 1-dimensional subspaces of T. Hence it is $AG(n, K)$, for some n. This completes the proof. □

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

14.8.2 Corollary. *A projective geometry of rank at least four is isomorphic to the geometry formed by the 1- and 2-dimensional subspaces of a vector space over a skew field.* □

Similarly we have the following.

14.8.3 Corollary. *Any Desarguesian projective plane is isomorphic to the plane formed by 1- and 2-dimensional subspaces of a 3-dimensional vector space over a skew field.* □

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

14.9 $PGL(2, \mathbb{F})$ on a Line

The *projective line* over $\mathbb{F}$ consists of the 1-dimensional subspaces of $V(2, \mathbb{F})$, the vector space of dimension two over $\mathbb{F}$. if we extend $\mathbb{F}$ by an element ∞ then the line spanned by the vector

$$\begin{pmatrix} x \\ y \end{pmatrix}$$

has a slope x/y; this slope determines the line. This we means we can view the points of the line as elements of $\mathbb{F} \cup \infty$. As

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$$

we see that if $ad - bc \neq 0$ then the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ maps the line with slope $(ax+by)/(cx+dy)$. Thus it determines a permutation of $\mathbb{F} \cup \infty$, this permutation may be called a *linear fractional mapping*. Two invertible 2×2 matrices determine the same linear fractional mapping if and only if one is a non-zero scalar multiple of the other. The linear fractional mappings form the group $PGL(2,\mathbb{F})$.

The matrices that fix ∞ are those of the form

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix},$$

these corresponds to the permutations of $\mathbb{F}$ that send x to $ax + b$, where $a \neq 0$. These mappings for the 1-dimensional affine group $AGL(1,\mathbb{F})$. You may show that, given two ordered pairs of elements of $\mathbb{F}$, there is a unique linear fractional map with $c = 0$ and $d = 1$ that maps the first pair to the second. Thus $AGL(1,\mathbb{F})$ acts 2-transitively on the points of $\mathbb{F}$, more precisely it is *sharply 2-transitive* on $\mathbb{F}$.

Since the linear fractional map $x \mapsto x^{-1}$ is a permutation of $\mathbb{F} \cap \infty$ that swaps 0 and ∞, we see that $PGL(2,\mathbb{F})$ acts transitively on $\mathbb{F} \cup \infty$, and it follows that it acts sharply 3-transitively. One consequence of this is that

$$|PGL(2,\mathbb{F})| = (q+1)q(q-1).$$

We have seen that $PGL(2,\mathbb{F})$ acts transitively on the ordered triples of distinct points from $\mathbb{F} \cup \infty$. We will consider its action on 4-tuples of distinct points.

If a, b, c, d are four elements of $\mathbb{F} \cup \infty$, their *cross ratio* $(a, b; c, d)$ is given by

$$(a, b; c, d) := \frac{(a-c)(b-d)}{(a-d)(b-c)}$$

14.9.1 Theorem. *Two 4-tuples of distinct elements of $\mathbb{F} \cup \infty$ are in the same orbit under $PGL(2,\mathbb{F})$ if and only if they have the same cross-ratio.* □

We leave the proof as an exercise. Note that to show that $PGL(2,\mathbb{F})$ preserves cross-ratio it is enough to show that cross-ratio is preserved by the affine maps and the inversion $x \mapsto x^{-1}$. (Although you need to supply an argument proving that these maps generate $PGL(2,\mathbb{F})$.)

If we choose to represent the points on the projective line by vectors in $\mathbb{F}^2$, then the cross-ratio of the four vectors a, b, c, d is equal to

$$\frac{[a,c][b,d]}{[a,d][b,c]}.$$

Here $[a,c]$ denotes the determinant of the 2×2 matrix with columns a and c. More generally if a,b,c,d are four points on a line ℓ and $p\notin\ell$ then the cross-ratio of these four points is equal to

$$\frac{[p,a,c][p,b,d]}{[p,a,d][p,b,c]}$$

You should verify this claim. Also show that if m is a second line then the cross-ratio of the points

$$pa\cap m,\ pb\cap m,\ pc\cap m,\ pd\cap m.$$

is equal to $(a,b;c,d)$. This provides with a way to define the cross-ratio of four concurrent lines.

14.10 Baer Subplanes

A set S of points and lines from an incidence structure is *dense* if each point and block of the structure is incident with an element of S. For example, if S consists of all points on some line and all lines on some point, then S is dense.

The fixed points of a collineation form a possibly degenerate projective geometry. So if the collineation is not a perspectivity, it fixes a projective geometry. A dense subgeometry is known as a *Baer subgeometry* and usually it will be a Baer subplane of a projective plane. If a projective plane contains a Baer subplane, the order of the plane must be a square. (This is easy to see for planes over a field, but it is true in general.)

14.10.1 Lemma. *The fixed points of an involutory collineation of a projective space form a dense subgeometry.* □

Exercises

1. Let $\mathbb{F}$ be a finite field of characteristic p. Show that any p-element in $PGL(n+1, \mathbb{F})$ fixes some point in $\mathbb{P}(n, \mathbb{F})$. Hence show that any Sylow p-subgroup of $PGL(n+1, \mathbb{F})$ fixes a maximal flag. Finally show that any maximal flag is fixed by a unique Sylow p-subgroup.

2. Let V the vector space of dimension d over $GF(q)$. If a, h are non-zero vectors in V, define the map $\tau_{h,a}$ from V to itself by

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

   If $h^T a \neq 1$, show that $\tau_{h,a}$ is a perspectivity of the projective space $\mathcal{P}(V)$ and determine its center and axis. Prove that $\mathcal{P}(V)$ is (p, H)-transitive for all p and H.

3. If a projective plane is (p, ℓ)-transitive and (q, ℓ)-transitive for distinct point p and q off ℓ, show that it is (r, ℓ)-transitive for all points r on $p \vee q$.

4. Let $\mathcal{G}$ be a projective geometry with rank at least four. Show that any two triangles in perspective from a point of $\mathcal{G}$ are in perspective from a line. (Note: if the triangles are coplanar, this will follow from a result in the notes, but they need not be coplanar and so there is still work to do.)

5. Show that a geometric partition T_i $(i = 1, \ldots, m)$ of an elementary abelian group T determines an affine geometry. HINT: first show that two disjoint lines are coplanar if and only if they are cosets of the same subgroup T_i. Then verify that the axioms for an affine geometry hold.

6. Let T be a group and let T_i, $(i = 1, \ldots, m)$ be a collection of at least two subgroups such that the sets $T_i \setminus 1$ partition T and, if $i \neq j$ then $T_i T_j = T$. Show that the incidence structure with the elements of T as its points and the right cosets of the subgroups T_i as its lines is an affine plane. (Note that we are not assuming that T is abelian, nor that the subgroups T_i are normal.)

7. Show that all finite translation planes have prime-power order, and that a translation plane of prime order is Desarguesian.

8. If α and β are two 4-tuples of points on the projective line over $\mathbb{F}$, show that there is an element of $PGL(2,\mathbb{F})$ that maps α to β if and only if their cross-ratios are equal.

9. If $\mathbb{E}$ is a quadratic extension of $\mathbb{F}$, the images under $PGL(2,\mathbb{E})$ of the points in $\mathbb{F} \cup \infty$ form a 3-$(q^2+1, q+1, 1)$ design. [A so-called Möbius plane.]

10. If a, b, c, d are four points on a line ℓ in the projective plane over $\mathbb{F}$ and p is a point off ℓ, prove that the cross-ratio of the four points is

$$\frac{[p,a,c][p,b,d]}{[p,a,d][p,b,c]}.$$

(Here $[p,a,c]$ is the determinant of the matrix with columns p, a, c, etc.)

11. Suppose a, b, c, d are four points on a line ℓ in the projective plane over $\mathbb{F}$ and p is a point not on ℓ. If m is a line not on p, show that the cross-ratio of the four points

$$m \cap pa,\ m \cap pb,\ m \cap pc,\ m \cap pd$$

is equal to $(a,b;c,d)$. [This effectively assigns a cross-ratio to 4-tuples of concurrent lines.]

12. If a projective plane of order n contains a Baer subplane, show that n is a square.

Chapter 15

Spreads and Planes

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

15.1 Spreads

Every projective geometry which is (p, H)-transitive for all points p on some hyperplane H gives rise, as we have seen, to a geometric partition of an abelian group T. The ring of endomorphisms of this partition is a skew field K and T is a vector space over K and the subgroups $T(p)$ are subspaces.

15.1.1 Lemma. *All components of a geometric partition have the same dimension as subspaces (over the kernel).*

Proof. Since $T(p)T(q)$ contains elements not in $T(p) \cup T(q)$ there is a point r, not equal to p or q, such that $T(r) \subseteq T(p)T(q)$. Since $T(p)T(r) = T(q)T(r)$ and $T(p)$, $T(q)$ and $T(r)$ are disjoint, it follows that $T(p)$ and $T(q)$ must have the same dimension. □

It is not hard to see that the geometry determined by the partition is a plane if and only if $T = T(p)T(q)$ for any pair of distinct points p and q. Since projective geometries with rank at least four are all of the form $\mathbb{P}(n, \mathbb{F})$, we no longer have much reason to bother working with geometric partitions in general. However spreads remain objects of considerable interest.

A *spread* is a collection of pairwise-skew m-dimensional subspaces of a $2m$-dimensional vector space V over a field $\mathbb{F}$ that partition the non-

zero vectors in V. A *partial spread* is a set of pairwise-skew subspaces of dimension m in V.

15.2 Coordinatizing Spreads

Two m-dimensional subspaces of V are skew if and only if the union of a basis from each subspace is a basis for V. We will present our subspaces as column spaces of $2m \times m$ matrices with rank m. In particular we define subspaces $V(\infty)$ and $V(0)$ as the respective column spaces of the matrices

$$\begin{pmatrix} 0 \\ I \end{pmatrix}, \quad \begin{pmatrix} I \\ 0 \end{pmatrix}.$$

Now suppose Y and Z respectively are the column spaces of the $2m \times m$ matrices

$$\begin{pmatrix} A \\ B \end{pmatrix}, \quad \begin{pmatrix} C \\ D \end{pmatrix},$$

each of rank m. Then Y and Z are skew if and only if the columns of

$$\begin{pmatrix} A & C \\ B & D \end{pmatrix}$$

are linearly independent, or equivalently if its determinant is not zero. We see that Y is skew to $V(\infty)$ if and only if A is invertible, and is skew to $V(\infty)$ if and only if B is invertible. In this case the matrices

$$\begin{pmatrix} A \\ B \end{pmatrix}, \quad \begin{pmatrix} I \\ BA^{-1} \end{pmatrix}$$

have the same column space. We have proved the following.

15.2.1 Lemma. *If Y is an m-dimensional subspace of V skew to $V(\infty)$ and $V(0)$, then Y is the column space of a matrix*

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

where A is invertible. □

If will be convenient to use $V(A)$ to denote the column space of the matrix in the above lemma. We note that $V(A)$ and $V(B)$ are skew if and only if $B - A$ is invertible.

15.2.2 Lemma. *If A is invertible there is a linear mapping of V that fixes $V(\infty)$ and $V(0)$ and maps $V(A)$ to $V(I)$.*

Proof. Exercise. □

So to each collection of pairwise skew subspaces of dimension n in V that contains $V(\infty)$ and $V(0)$, there is a set S of invertible matrices such that the difference of any two matrices in S is invertible. The set $S \cup 0$ has the property that the difference of any two matrices in it is invertible. If the difference of two matrices M_1 and M_2 is invertible then the first row of $M_1 - M_2$ cannot be zero. Hence a set of $m \times m$ matrices over $GF(q)$ such that the difference of any two distinct matrices is invertible has size at most q^m, and set of $m \times m$ matrices such that the difference of any two distinct matrices is invertible has size at most q^m. If $\mathcal{M}$ is a spread set and $M \in \mathcal{M}$, then the set

$$\{N - M : N \in \mathcal{M}\}$$

is a set of $m \times m$ matrices, one of which is 0, such that the difference of any two distinct matrices is invertible. We call such a set a *spread set*. The subspaces associated to a spread set, along with $V(\infty)$, provides us with $q^m + 1$ pairwise-skew subspaces of dimension m, each of which contains $q^m - 1$ non-zero vectors. Since

$$|V| - 1 = q^{2m} - 1 = (q^m + 1)(q^m - 1),$$

each spread set determines a spread in V.

15.2.3 Lemma. *If Σ is a spread set of invertible $m \times m$ matrices over $\mathbb{F}$ and u is a non-zero vector in $U = \mathbb{F}^m$, then for each non-zero vector v in U there is a unique element M of Σ such that $Mu = v$.* □

Suppose Σ is a spread set of invertible $d \times d$ matrices over $\mathbb{F}$, let U be $\mathbb{F}^d$ and let V be the direct sum $U \oplus U$. Let $\mathcal{A}$ be the affine plane determined by the spread. We can identify the points of $\mathcal{A}$ with the vectors in V, which write in the form (x, y) where x and y each have length d. We define $V(\infty)$ to be the subspace formed by the vectors $(0, y)$, and if $\sigma \in \Sigma \cup 0$ then

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

The lines of $\mathcal{A}$ are the cosets of the subspace $V(\sigma)$, for $\sigma \in \Sigma \cup \{0, \infty\}$. (So we may view the elements of Σ are finite non-zero slopes. It may also help to think of $V(\infty)$ as the y-axis and $V(0)$ as the x-axis.)

15.3 The Complex Affine Plane

Define the set $\mathcal{M}$ of 2×2 real matrices by

$$\mathcal{M} = \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}.$$

This a vector space over $\mathbb{R}$. Since

$$\det\left(\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \right) = a^2 + b^2,$$

we see that all non-zero elements of $\mathcal{M}$ are invertible. Hence $\mathcal{M}$ determines a spread in $\mathbb{R}^4$.

The affine plane we obtain is isomorphic to the usual affine plane over $\mathbb{C}$. As an exercise you might show that the kernel of this spread is isomorphic to $\mathbb{C}$.

We can generalize this example. Let $\mathcal{E}$ be an extension of the field $\mathbb{F}$ and assume $\mathcal{E}$ has dimension d over $\mathbb{F}$. Then multiplication by an element of $\mathcal{E}$ is a linear mapping of $\mathcal{E}$, and so it can be represented by a $d \times d$ matrix over $\mathbb{F}$. Thus the space of $d \times d$ matrices over $\mathbb{F}$ contains a subspace consisting of matrices representing elements of $\mathcal{E}$. Any non-zero element of this subspace is invertible, and thus these form a spread set of invertible matrices. The resulting plane is the usual plane over $\mathcal{E}$.

15.4 Collineations of Translation Planes

We prove what might be called the fundamental theorem for translation planes. Note that any collineation of a translation plane can be expressed as a product of a translation and an automorphism of the plane that fixes o, so this result is providing more information that might be evident at first glance.

15.4.1 Theorem. *Let $\mathcal{A}$ be the affine plane determined by the spread $\mathcal{S}$ of $V(2n, \mathbb{F})$ and let K be its kernel. Then the collineations of $\mathcal{A}$ which fix 0 are induced by the semilinear mappings of V which map the components of $\mathcal{S}$ onto themselves.*

Proof. Let α be a collineation of $\mathcal{A}$ fixing 0. If $v \in V$, the mapping $\tau(u)$ on the points of $\mathcal{A}$ by

$$x^\tau(u) = x + u$$

is a translation and the mapping $u \mapsto \tau(u)$ is an isomorphism from V to the group of translations of $\mathcal{A}$.

If τ is a translation then it is easy to see that $\alpha^{-1}\tau\alpha$ is perspectivity with axis the line at infinity. Since $\alpha^{-1}\tau\alpha$ fixes any line on the centre of τ, it follows that it is a translation. Hence the mapping

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

is an automorphism of the group of translations of $\mathcal{A}$ and therefore it induces an additive mapping of V.

Any homology of $\mathcal{A}$ with centre 0 and axis the line at infinity maps v in V to κv, for some κ in K. Then $\alpha^{-1}\kappa\alpha$ is a homology with centre 0 and axis ℓ_∞, hence corresponds to an element κ^a (say) of K. Now

$$(v^\kappa)^\alpha = (v^\alpha)^{\alpha^{-1}\kappa\alpha} = (v^\alpha)^{\kappa^a}$$

and, if $\lambda \in K$, this implies that

$$\begin{aligned} v^{\alpha(\kappa+\lambda)^a} = (v^{\kappa+\lambda})^\alpha &= \left(v^\kappa + v^\lambda\right)^\alpha \\ &= v^{\kappa\alpha} + v^{\lambda\alpha} \\ &= v^{\alpha\kappa^a} + v^{\alpha\lambda^a} \\ &= v^{\alpha(\kappa^a+\lambda^a)}. \end{aligned}$$

So $(\kappa + \lambda)^a = \kappa^a + \lambda^a$. Further,

$$v^{\alpha(\kappa\lambda)^a} = \left(v^{\kappa\lambda}\right)^a = v^{\kappa a\lambda^a} = v^{\alpha\kappa^a\lambda^a}$$

and thus $(\kappa\lambda)^a = \kappa^a\lambda^a$. This shows that the map

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

is an automorphism of K. We conclude that α is semilinear. □

Let $V = U \oplus U$ and assume Σ is a spread set of invertible matrices. The elements of Σ belong to $GL(U)$. The spread set determines a spread in V that contains $V(0)$ and $V(\infty)$; let $\mathcal{A}$ be the corresponding affine plane. We write the elements of V in the form (u, v), where $u, v \in U$. If $\sigma \in \Sigma$, then

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

and

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

The subspaces $V(\sigma)$ are the lines through the point $(0, 0)$ in $\mathcal{A}$.

Since U is a vector space over the kernel K of $\mathcal{S}$, it follows that any non-identity automorphism of K must act non-trivially on it. (That is, it cannot fix each element of U.) From this it follows in turn that any perspectivity of $\mathcal{A}$ fixing 0 must be induced by a linear mapping of V, and not just a semilinear one.

If $V = U \oplus U$ then any linear M mapping of V can be written in partitioned form

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where the blocks are $d \times d$ matrices. If M fixes $V(0)$ then

$$\begin{pmatrix} I \\ 0 \end{pmatrix}, \quad \begin{pmatrix} A \\ C \end{pmatrix}$$

have the same column space; therefore $C = 0$ and A is invertible. If M fixes $V(\infty)$ we see similarly that $B = 0$ and D is invertible. If $1 \in \Sigma$ and M fixes $V(1)$, then the column spaces of the matrices

$$\begin{pmatrix} I \\ I \end{pmatrix}, \quad \begin{pmatrix} A \\ D \end{pmatrix}$$

are equal and consequently $A = D$. If M fixes each component of the spread given by Σ, then

$$\begin{pmatrix} I \\ \sigma \end{pmatrix}, \quad \begin{pmatrix} A \\ A\sigma \end{pmatrix}$$

have the same column space and so $A\sigma A^{-1} = \sigma$, for each σ in Σ. In other terms, A commutes with each element of Σ. As an exercise, show that any non-zero matrix that commutes with each element of a spread set is invertible.

15.5 The Fundamental Theorem

15.4.1 holds for any Desarguesian affine space, we just work with geometric partitions rather than spreads. It leads to a proof of the fundamental theorem of projective geometry, which we outline here.

15.5.1 Theorem. *Any collineation of a Desarguesian projective space is the composition of an invertible linear map and a field automorphism.*

Proof. Suppose α is a collineation of a Desarguesian projective space. If α fixes a point it fixes a hyperplane and, if the fixed point is not on the hyperplane, then α is semilinear. So suppose α does not fix a point. If p is a point, we can find a homology γ such that $p^{\alpha\gamma} = p$. Then $\alpha\gamma$ fixes a hyperplane H and if $p \notin H$ then $\alpha\gamma$ is semilinear. Since γ is semilinear it follows that α is too. Now suppose our space has rank d and the points $x_1, \ldots, x_d$ are a basis. Then there is a collineation γ, a semilinear mapping, such that x_i is fixed by $\alpha\gamma^{-1}$ for each i. Then $\alpha\gamma^{-1}$ fixes the hyperplane spanned by $x_1, \ldots, x_{d-1}$ and fixes the point x_d, which is not on this hyperplane. Therefore it is semilinear. □

15.6 Collineations and Spread Sets

We relate collineations with properties of spread sets. We consider the line at infinity ℓ_∞ in $\mathcal{A}$ as a distinguished line, rather than as a missing line. Let (0) and (∞) be the points at which $V(0)$ and $V(\infty)$ respectively meet ℓ_∞.

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

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

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

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

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

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

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

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

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

The converse to this corollary is false. (See the next section.) There is an analog of 15.6.1 for elations:

15.6.3 Lemma. *Let Σ_0 denote $\Sigma \cup 0$. The set*

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

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

Proof. If α is represented by the matrix

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

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

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

If $(0, v) \in V(\infty)$ then $(0, v)\alpha = (Yv, Zv)$. Hence if each point on $V(\infty)$ is fixed by α then $Y = 0$ and $Z = I$. If α also fixes all lines through (∞) then

it must fix the cosets of $V(\infty)$. The elements of a typical coset of $V(\infty)$ have the form $(a, b+v)$, where v ranges over the elements of U. Now

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

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

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

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

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

Proof. Since $\mathcal{P}$ is (p, l)-transitive for all points p and lines l, we deduce from 15.6.1 that Σ is a group under multiplication and from Lemma **??** that Σ_0 is a group under addition. If $\sigma \in \Sigma$ then $\sigma^{-1}\Sigma = \Sigma$, implying that $I = \sigma^{-1}\sigma \in \Sigma$. As both addition and multiplication are the standard matrix operations, the usual associative and distributive laws hold. Therefore Σ_0 is a skew field. □

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

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

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

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

15.7 A Nearfield Plane

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

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

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

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

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

15.8 The Lorimer-Rahilly Plane

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

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

Each line gives us two 3-cycles belonging to the alternating group A_7. (For example the line 257 produces (257) and (275).) Let Σ be the set formed

by these fourteen 3-cycles, together with the identity. Let X be the 4-dimensional vector space over $GF(2)$. We claim that A_7 can be viewed as a subgroup of $GL(4,2)$ acting transitively on the 15 non-zero elements of X. (The proof of this is given in the next section.)

We prove that if σ and τ are elements of Σ then $\sigma^{-1}\tau$ acts fixed-point freely on the non-zero vectors of X. A routine check shows that $\sigma^{-1}\tau$ is either a 3-cycle or a 5-cycle. If x is a non-zero vector in X then the subgroup of A_8 leaving it fixed has order $8!/30 = 21 \cdot 2^6$. Thus this subgroup contains no elements of order 5, and so all elements of order 5 in A_8 must act fixed-point freely. Suppose then that $\theta = \sigma^{-1}\tau$ is 3-cycle in A_8. Then there is a 5-cycle ϕ which commutes with θ. If x is non-zero vector fixed by θ then

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

and so $x\phi$ is also fixed by θ. This shows that the number of non-zero vectors fixed by θ is divisible by 5. As θ has order three, the number of non-zero vectors not fixed by it is divisible by three. This implies that θ cannot fix 5 or 10 vectors, and hence that it must have 15 fixed points, that is, it is the identity element.

Thus we have shown that Σ determines a spread of $X \oplus X$. The resulting plane is not a nearfield plane, for then Σ would be a group of order 15. The only group of order 15 is cyclic, and hence abelian. But the Sylow 2-subgroups of $SL(2,3)$ are isomorphic to the quaternion group, which is not abelian. The plane we have constructed is called the *Lorimer-Rahilly* plane. Note that the collineation group of the plane over $GF(2)$ induces a group of collineations of the new plane fixing $V(0)$, $V(1)$ and $V(\infty)$, and acting transitively on the remaining components.

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

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

plane of order two, but 'heptad' is shorter) We say that a set of triples are *concurrent* if there is some point common to them all, and the intersection of any two of them is this common point. A *star* is a set of three concurrent triples. The remainder of the argument is broken up into a number of separate claims.

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

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

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

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

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

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

15.9.3 Claim. *There are exactly 30 heptads.*

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

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

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

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

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

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

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

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

Simple counting.

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

More counting.

Now we construct a linear space. Choose one orbit of heptads under the action of A_7, and call its elements points. Let the triples be the lines, and say that a point is on a line if the correponding heptad contains the triple. The elements of the second orbit of heptads under A_7 determine subspaces of rank three, each isomorphic to a projective plane. It is now an exercise to show that there are no other non-trivial subspaces, and thus we have a

linear space of rank four, with all subspaces of rank three being projective planes. Hence our linear space is a projective geometry, of rank four. Since its lines all have cardinality three, it must be the projective space of rank four over $GF(2)$. As $GF(2)$ has no automorphisms, the collineation group of our linear space consists entirely of linear mappings; hence it is isomorphic to $GL(4,2)$. (Note that we have just used the characterization of projective geometries as linear spaces with all subspaces of rank three being projective planes, the fact that projective geometries of rank at least four are all of the form $\mathbb{P}(n,\mathbb{F})$ and the fundamental theorem of projective geometry, i.e., that the collineations of $\mathbb{P}(n,\mathbb{F})$ are semilinear mappings.) Our argument has thus revealed that A_7 is isomorphic to a subgroup of $GL(4,2)$. A direct computation reveals that it has index eight.

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

15.10 Moufang Planes

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

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

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

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

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

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

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

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

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

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

$$\rho_a(x) = xa$$

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

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

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

Exercises

1. Let Σ be subset of $GL(U)$ determining a spread π of $V = U \oplus U$. If $\sigma \in \Sigma$ and $\sigma\Sigma^{-1}\sigma = \Sigma$, show that the mapping which sends (u, v) to $(v\sigma^{-1}, u\sigma)$ is a perspectivity of the plane $\mathcal{P}(\pi)$ with axis $V(\sigma)$ which interchanges $V(0)$ and $V(\infty)$. From this deduce that the collineation group of a nearfield plane has at most three orbits on its points, and no fixed points.

2. Let $\mathcal{P}^\ell$ be a nearfield plane. Suppose that m is a translation line in $\mathcal{P}$ distinct from ℓ. The group of all perspectivities with axis m and centre $\ell \cap m$ acts transitively on the points of $\ell \setminus m$. Deduce from this that group of collineations of $\mathcal{P}$ fixing ℓ acts transitively on the points of ℓ, and hence that there is a point y of $\ell \setminus m$ and a line h on $\ell \cap m$ not containing y such that $\mathcal{P}$ is (y, h)-transitive. Finally deduce that $\mathcal{P}$ is Desarguesian.

3. Use the results of the previous two exercises to show that a non-Desarguesian nearfield plane is not self-dual.

4. Let U be a vector space and let Σ be a subset of $GL(U)$ determining a spread of $V = U \oplus U$. If Σ is a group, show that the mappings $t(\sigma, u)$ with $\sigma \in \Sigma$ and $u \in U$ given by

$$t(\sigma, u) : x \mapsto x\sigma + u$$

   form a sharply 2-transitive group of permutations of U. (This shows that every nearfield plane determines a sharply 2-transitive permutation group. It is not too difficult to show that every sharply 2-transitive group determines an affine plane and with more effort it can be shown that the resulting planes are translation planes.)

5. Show that if the dual of $\mathcal{A}$ is a translation plane, then there is a spread set Σ for $\mathcal{A}$ such that $\Sigma \cup 0$ is an additive group.

6. Show that any non-zero matrix that commutes with each element of a spread set is invertible.

7. Let π be a spread with affine plane $\mathcal{A}$. Let Σ be the corresponding spread set, and assume $1 \in \Sigma$. If $\sigma \in \Sigma$, show that the map

$$(x, y) \mapsto (y^{\sigma^{-1}}, x^\sigma)$$

a perspectivity with axis $V(\sigma)$ that swaps (0) and (∞) if and only if $\sigma\Sigma^{-1}\sigma = \Sigma$. [Check]

Chapter 16

Varieties

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

16.1 Definitions

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

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

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

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

The affine variety determined by a homogeneous polynomial g is said to be a *cone*. More generally, $\mathcal{V}_g$ is a cone at a point $\mathbf{a}$ if $g(\mathbf{y})$ is a homogeneous polynomial in $\mathbf{y} = \mathbf{x} - \mathbf{a}$. The projective variety associated with $\mathcal{V}_g$ is also said to be a cone at $\mathbf{a}$.

16.2 Is There a Point?

We present a result that we can use to show that the zero set of a homogeneous polynomial is not empty.

16.2.1 Theorem. *Let f be a polynomial of degree k in n variables over*

the field $\mathbb{F}$ with $q = p^r$ elements. If $k < n$ then the number of solutions of $f(\mathbf{x}) = 0$ is zero modulo p.

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

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

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

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

$$\sum_{\mathbf{a} \in \mathbb{F}^n} f(\mathbf{a})^{q-1}. \tag{16.2.1}$$

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

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

where

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

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

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

is congruent to zero modulo p. This implies in turn that (16.2.1) is congruent to zero modulo p. □

The following result is due to Chevalley.

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

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

16.3 The Tangent Space

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

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

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

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

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

16.4 Tangent Lines

If f is a homogeneous polynomial then the *degree* of the hypersurface $\mathcal{V}_f$ is the degree of f. The degree is important because it is an upper bound on

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

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

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

$$f(\mathbf{a} + t\mathbf{b}) = F^{(0)}(\mathbf{b}) + tF^{(1)}(\mathbf{b}) + \cdots + t^k F^{(k)}(\mathbf{b}), \tag{16.4.1}$$

where $F^{(i)}$ is a polynomial in the entries of $\mathbf{b}$, with coefficients depending on $\mathbf{a}$.. If the first nonzero term in (16.4.1) has degree m in t, we say that the *intersection multiplicity* at a of $a \vee b$ and $\mathcal{V}_f$ is m. If $a \in \mathcal{V}$ then $F^{(0)}(\mathbf{b}) = 0$; thus the intersection multiplicity is greater than zero if and only if a is on $\mathcal{V}$. We have

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

and so the intersection multiplicity is greater than 1 if and only if b lies in the tangent space $T_a(f)$. Since $a \in T_a(f)$, the point b is in $T_a(f)$ if and only if the line $a \vee b$ lies in $T_a(f)$. A line having intersection multiplicity greater than one with $\mathcal{V}_f$ is a *tangent line*. We have just shown that $T_a(f)$ is the union of all the tangent lines to $\mathcal{V}_f$ at a. A subspace is *tangent* to $\mathcal{V}_f$ at a if it is contained in $T_a(f)$. It is possible for the hypersurface $\mathcal{V}$ to completely contain a given line $a \vee b$. In this case the left side of (16.4.1) must be zero for all t, whence it follows that $a \vee b$ is a tangent. More generally, a subspace contained in $\mathcal{V}_f$ is tangent to $\mathcal{V}_f$ at each point in it. There is another important consequence of (16.4.1) which must be remarked on.

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

Proof. Let l be a line a let a be a point on l which is not on the hypersurface $\mathcal{V}_f$. The points of l on $\mathcal{V}$ are given by the solutions of (16.4.1). If this is identically zero then l is contained in the hypersurface, otherwise it is polynomial of degree k and has at most k zeros. □

We have only considered tangent spaces to hypersurfaces. Everything extends nicely to the case of varieties; we simply define the tangent space of the variety $\mathcal{V}$ at a to be the intersection of the tangent spaces at a of the hypersurfaces which intersect to form $\mathcal{V}$. Since we will not be working with tangent spaces to anything other than hypersurfaces, we say no more on this topic.

16.5 Tangents to Quadrics

The tangent space to a quadric is easily described, using the following result, which is a special case of Taylor's theorem.

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

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

The matrix $H(f)$ is the *Hessian* of f. It is a symmetric matrix and, if the characteristic of $\mathbb{F}$ is even, its diagonal entries are zero. The tangent plane at the point $\mathbf{a}$ has equation $\mathbf{a}^T H \mathbf{x} = 0$, and therefore $\mathbf{a}$ is singular if and only $\mathbf{a}^T H = 0$. Consequently the quadric determined by f is smooth if and only if no point of the quadric lies in the kernel of H. Obviously a sufficient condition for this is that $H(f)$ is non-singular. (However it is possible for the quadric to be smooth when H is singular. For example, consider any smooth conic in a projective plane over a field of even order.)

If the characteristic of $\mathbb{F}$ is not even then $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H(f)\mathbf{x}$. Since we do not wish to restrict the characteristic of our fields, we will not be making use of this observation. One important consequence of ??Lemma 3.2 is that if a tangent to a quadric at a meets it at a second point b then it is contained in the quadric. (For these conditions imply that $f(\mathbf{a}) = \mathbf{a}^T H \mathbf{b} = f(\mathbf{b}) = 0$.)

Since all lines through a singular point are tangents, it follows that a line which passes through a singular point and one other point must be contained in the quadric. Of course any line meeting a quadric in three or more points must be contained in it, by ??Lemma 3.1. A line which meets a quadric in two points is a *secant*.

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

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

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

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

16.5.3 Lemma. *If a is a singular point on a quadric $\mathcal{Q}$ then all lines through a are tangents. The singular points on $\mathcal{Q}$ form a subspace.*

Proof. If $a \in \mathcal{Q}$ then

$$f(\lambda a + \mu x) = \lambda\mu a^T H x + \mu^2 x$$

and therefore if a is singular, then either $f(x) \neq 0$ and $a \vee x$ meets $\mathcal{Q}$ in a, or $f(x) = 0$ and $a \vee x$ is contained in Q.

For the second claim, if a and b are singular points then $(\lambda a + \mu b)^T H = 0$ for all λ and μ, and so all points points on $a \vee b$ are singular. □

If a and b are distinct singular points on $\mathcal{Q}$, then the first claim implies

16.6 Intersections of Hyperplanes and Hypersurfaces

Suppose that f is a homogeneous polynomial defined over a field $\mathbb{F}$. Then f is irreducible if it does not factor over $\mathbb{F}$, and it is *absolutely irreducible* if it does not factor over the algebraic closure of $\overline{\mathbb{F}}$ of $\mathbb{F}$. If g is a factor of

F over $\overline{\mathbb{F}}$ then $\mathcal{V}_g$ is a *component* of $\mathcal{V}_f$. Thus $\mathcal{V}_f$ is a union of components, although not necessarily a disjoint union. Over finite fields the situation is a little delicate, in that $\mathcal{V}_g$ may be empty. However this possibility will not be the source of problems—such components tend to remain completely invisible.

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

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

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

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

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

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

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

$$\lambda f + \mu g. \tag{16.6.1}$$

Each of these is a quadratic, and thus describes a conic. If x is a point not on the 4-arc then the member of (16.6.1) with $\lambda = g(\mathbf{x})$ and $\mu = -f(\mathbf{x})$ vanishes at x and at each point of the 4-arc. This establishes the existence of a conic through any set five points containing a 4-arc. Suppose now that $\mathcal{C}$ and $\mathcal{C}'$ are two conics meeting on the 4-arc *abcd* and the fifth point p. By Bézout's lemma, these two conics must have a common component. If the conics are distinct, this component must be described by a linear polynomial, i.e., it must be a line ℓ. Hence $\mathcal{C}$ and $\mathcal{C}'$ must each be the union of two lines, possibly the same line twice. But now each conic contains ℓ and at least two points from the 4-arc not on ℓ. We conclude that the conics must coincide. □

The hypersurfaces determined by the set of polynomials

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

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

Exercises

1. Let f_α ($\alpha \in A$) be a set of polynomials in n variables over $\mathbb{F}$. Prove that if f_α has degree k_α and $\sum_\alpha k_\alpha < n$, the number of common zeros of the polynomials is congruent to zero modulo p. (*Hint:* use the function $\prod_\alpha (1 - f_\alpha^{q-1})$).

2. If g and h are polynomials in $n+1$ variables and gh is homogeneous, show that g and h are homogeneous. (You can look this up somewhere, if you like.)

3. Show that the pencil of conics through a 4-arc in $\mathbb{P}(2, \mathbb{F})$ contains exactly three singular conics.

4. If $\mathcal{V}$ is a hypersurface of degree at least 2 over an algebraically closed field, show that any line which meets it in exactly one point is a tangent.

Chapter 17

Conics

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

17.1 The Kinds of Conics

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

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

(a) a single point,

(b) a double line,

(c) the union of two distinct lines, or

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

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

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

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

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

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

We now prove the theorem. Suppose that $\mathcal{C}$ is a conic. Assume first that it contains two singular points a and b. By Lemma 16.5.3 all points on $a \vee b$ must belong to $\mathcal{C}$. If c is a point of the conic not on $a \vee b$ then all points on $c \vee a$ and $c \vee b$ must also lie in $\mathcal{C}$. If x is a point in $\mathbb{P}(2, \mathbb{F})$ then there is a line through x meeting $c \vee a$, $c \vee b$ and $a \vee b$ in distinct points. Hence this line lies in $\mathcal{C}$ and so $x \in \mathcal{C}$. This proves that $\mathcal{C}$ is the entire plane, which is impossible. Thus we have shown that if $\mathcal{C}$ contains two singular points then it must consist of all points on the line joining them, i.e., it is a repeated line.

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

$\mathcal{C}$ must lie in a line contained in $\mathcal{C}$. Hence $\mathcal{C}$ must contain two distinct lines, and their point of intersection is singular. Consequently $\mathcal{C}$ can contain no lines, but must rather be a $(q+1)$-arc. □

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

17.2 Conics and Cross-Ratio

We start by reconsidering conics through five points. Suppose the vectors a, b, c, d in $\mathbb{F}^3$ form a 4-arc. If x, y, z are three vectors in $\mathbb{F}^3$, we use $[x, y, z]$ to denote the determinant of the matrix with columns x, y and z.

The x represents a point on $a \vee b$ if and only $[x, a, b] = 0$, and so the equation of the conic formed by the lines $a \vee b$ and $c \vee d$ is

$$[x, a, b][x, c, d] = 0.$$

Similarly the equation of the conic formed from $a \vee d$ and $b \vee c$ is

$$[x, a, d][x, b, c]$$

and hence any conic in the pencil spanned by these two conic has the form

$$\lambda[x, a, b][x, c, d] + \mu[x, a, d][x, b, c].$$

If we choose a fifth point e, then e is on this conic if

$$\frac{\lambda}{\mu} = -\frac{[e, a, d][e, b, c]}{[e, a, b][e, c, d]}$$

and so the equation for the conic on a, b, c, d, e is

$$[e, a, d][e, b, c][x, a, b][x, c, d] - [e, a, b][e, c, d][x, a, d][x, b, c] = 0.$$

We can rewrite this equation as

$$\frac{[x, a, b][x, c, d]}{[x, a, d][x, b, c]} = \frac{[e, a, b][e, c, d]}{[e, a, d][e, b, c]}.$$

Now we observe that each side is a cross-ratio—the left side of the cross-ratio of the four lines that join x to a, b, c, d, the right side is the cross-ratio for the four lines through e.

17.3 Pascal and Pappus

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

$$f(\mathbf{x}) = (\mathbf{x}^T\mathbf{a}_{01})(\mathbf{x}^T\mathbf{a}_{23})(\mathbf{x}^T\mathbf{a}_{45}) \tag{17.3.1}$$

is homogeneous with degree three. Similarly, the three sides opposite to those used in (17.3.1) determine a second cubic, g say. By Bézout's lemma, two cubics with no common component meet in at most nine points. A common component of our two cubics would have to contain a line, and our hypothesis that the sides are distinct prevents this. Therefore $\mathcal{V}_f$ and $\mathcal{V}_g$ meet in the six points of our hexagon, together with the points of intersection of the three pairs of opposite sides.

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

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

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

lie on a line l, with equation $\mathbf{a}^T\mathbf{x} = 0$. Let f and g be the two cubics defined above. For any scalars λ and μ, the polynomial $F = \lambda f + \mu g$ is cubic and contains the nine points in which $\mathcal{V}_f$ and $\mathcal{V}_g$ intersect. We wish to choose the scalars so that the line l is contained in $\mathcal{V}_F$. If l has only three points, there is no work to be done. Thus we may choose a fourth point p on l, and choose λ and μ so that $F(p) = 0$. Thus the cubic curve $\mathcal{V}_F$ meets the line l in four points, and if we extend $\mathbb{F}$ to its algebraic closure, then the line extending l still meets the extension of $\mathcal{V}_F$ in at least four points. Bézout's theorem now implies that l must be contained in the curve and

so we deduce, by ??Lemma 4.4.2, that $F = (\mathbf{a}^T\mathbf{x})G$ for some polynomial F_1. But G must be homogeneous of degree two and therefore $\mathcal{V}_G$ is a conic. Thus $\mathcal{V}_F$ is the union of the line l and the conic $\mathcal{V}_G$. If the hexagon is contained in the union of two lines then it is on a conic, and we are finished. Otherwise a simple check shows that no points on the hexagon lie on L (do it), hence they line on the conic. This proves the first part of the theorem.

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

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

$$F(\mathbf{x}) := \lambda f(\mathbf{x}) + \mu g(\mathbf{x}) = 0 \tag{17.3.2}$$

in at least the six points of the hexagon. As $|\mathcal{C}| \geq 6$, our field must have order at least five. If it is exactly five then $\mathcal{C}$ is contained in the solution set of (17.3.2) for any choice of scalars; otherwise we may choose a point p of $\mathcal{C}$ not in the hexagon and then choose λ and μ so that $\mathcal{V}_F$ meets $\mathcal{C}$ in at least seven points. By Bézout's theorem, this implies that these two curves have a common component. The only component of $\mathcal{C}$ is $\mathcal{C}$ itself, thus $F = hG$ for some linear polynomial G. Hence $\mathcal{V}_F$ is the union of a line and the conic $\mathcal{C}$, and the points of intersection of the opposite sides of our hexagon must be on the line. □

Pappus' theorem is the assertion that the intersections of the opposite sides of a hexagon are collinear if the points of the hexagon lie on two lines. It is particularly important because it can be proved that a projective plane has the form $\mathbb{P}(2, \mathbb{F})$, where $\mathbb{F}$ is a field, if and only if Pappus' theorem holds. Thus, if we could prove geometrically that Pappus' theorem held in all finite

Desarguesian planes then we would have a geometric proof that a finite skew field is a field. No such proof is known. Planes for which Pappus' theorem is valid are called *Pappian*. All Pappian planes are, of course, Desarguesian.

17.4 Automorphisms of Conics

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

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

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

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

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

One immediate consequence of this theorem is the following result.

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

Proof. If $|\mathbb{F}| = 2$ or 3, this result can be verified easily. Assume then that $|\mathbb{F}| > 3$. From the theorem, $\mathrm{Aut}(\mathcal{C})$ is 2-transitive on the points of $\mathcal{C}$.

To prove that $\mathrm{Aut}(\mathcal{C})$ is 3-transitive it will suffice to prove that if A, B, C and D are four points on $\mathcal{C}$ then there is an automorphism of it fixing A and B and mapping C to D. Let X be a fifth point on the conic. By the theorem, there is an involution in $\mathrm{Aut}(\mathcal{C})$ swapping A and B, and sending

C to X. Similarly, there is an involution swapping B and A and sending X to D. The product of these two involutions is the required automorphism.

Next, suppose that A, B and C are three points on the conic. Any automorphism which fixes these three points must fix the tangents at A and B. Hence it fixes their point of intersection, which we denote by P. Thus the automorphism fixes each point in a 4-arc, and the only element of $PGL(3, \mathbb{F})$ which fixes a 4-arc is the identity. □

It follows at once from the corollary that if $|\mathbb{F}| = q$ then $|\mathrm{Aut}(\mathcal{C})| = q^3 - q$. We have already seen that the conics in $\mathbb{P}(2, \mathbb{F})$ correspond to the points in $\mathbb{P}(5, \mathbb{F})$, and are thus easily counted, there are

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

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

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

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

$$N(q+1)q(q-1)(q-2)(q-3) = n_5. \tag{17.4.1}$$

We find that

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

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

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

and, on comparing this with (17.4.1), we obtain that $N = (q^2+q+1)q^2(q-1)$ as claimed. □

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

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

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

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

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

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

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

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

17.5 Linear Mappings of Quadratic Polynomials

Linear fractional mappings act of the space of homogeneous polynomials over $\mathbb{F}$ with degree at most two: we have

$$p(x,y) \mapsto p(ax+by, cx+dy).$$

These polynomials can be divided into three classes, according as they have two distinct roots, one root with multiplicity two, or no roots over $\mathbb{F}$. It is a routine exercise to show that $PGL(2,\mathbb{F})$ acts transitively on the $q+1$ polynomials with one root of multiplicity two. This follows from the fact that $PGL(2,\mathbb{F})$ acts transitively on $\mathbb{F}\cup\infty$. But, of course, we know that it acts 3-transitively on this set. Hence it is 2-transitive, and we deduce that $PGL(2,\mathbb{F})$ acts transitively on the $\binom{q+1}{2}$ polynomials with two distinct roots. It remains to determine how many orbits of irreducible polynomials there are. Any such polynomial can be written in the form

$$(s+\omega t)(s+\bar{\omega}t), \tag{17.5.1}$$

where ω is an element of a fixed quadratic extension $\mathbb{F}(\theta)$ of $\mathbb{F}$, not contained in $\mathbb{F}$, and $\bar{\omega}$ is the image of ω under 'complex conjugation'. (If $\mathbb{F}$ is finite of order q then $\bar{\omega} = \omega^q$.) We prove that if $\rho \in \mathbb{F}(\theta) \setminus \mathbb{F}$ then there is an element of $PGL(2, \mathbb{F})$ mapping the polynomial with roots ω and $\bar{\omega}$ to the polynomial with roots ρ and $\bar{\rho}$. Our mapping α sends $s + \omega t$ to

$$as + bt + \omega(cs + dt) = (a + \omega c)s + (b + \omega d)t.$$

If $c = 0$ and $a = 1$ then $s + \omega t$ is mapped to $s + (b + \omega d)t$. The equations

$$b + \omega d = \rho, \quad b + \bar{\omega} d = \bar{\rho}$$

can be solved uniquely, yielding

$$b = \frac{\bar{\omega}\rho - \omega\bar{\rho}}{\bar{\omega} - \omega}, \quad d = \frac{\rho - \bar{\rho}}{\omega - \bar{\omega}}$$

As both solutions b and d lie in $\mathbb{F}$, it follows that $PGL(2, \mathbb{F})$ acts transitively on the irreducible polynomials of degree two.

17.6 Affine Conics

We have shown that $PGL(2, \mathbb{F})$ has three orbits on homogeneous polynomials of degree two in two variables. It remains to see what the geometric implications of this fact are.

Each homogeneous polynomial can be represented by the vector of its coefficients, in fact we have a bijection between polynomials of degree two and lines in $\mathbb{P}(2, \mathbb{F})$. Thus $as^2 + bst + ct^2$ corresponds to the line with coordinate vector (a, b, c). The points on the conic $\mathcal{C}$ with equation $x_1^2 - x_0 x_2$ all have the form $(\lambda^2, -\lambda, 1)^T$, without loss of generality. The line determined by a polynomial with a double root contains exactly one point on this conic, and is thus a tangent to it. The lines corresponding to the polynomial with two distinct roots determine secants to the conic, while the lines corresponding to the irreducible polynomial are external lines.

Hence we have shown that the group of collineations of $\mathbb{P}(2, \mathbb{F})$ induced by $PGL(2, \mathbb{F})$ fixes $\mathcal{C}$ and has three orbits on the lines of $\mathbb{P}(2, \mathbb{F})$. An immediate consequence of this is that the automorphism group of a smooth conic must be isomorphic to $PGL(2, \mathbb{F})$, since we have shown previously that $|\mathrm{Aut}(\mathcal{C})| = q^3 - q = |PGL(2, \mathbb{F})|$. Second, $\mathrm{Aut}(\mathcal{C})$ acts transitively on

the secants, tangents and external lines to $\mathcal{C}$. We can obtain a conic in the affine plane by choosing a line at infinity. We now know that there are only three different ways of doing this. The resulting conic is a hyperbola, parabola or ellipse, according as the line at infinity meets the projective conic in two, one or zero points.

17.7 Ovals

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

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

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

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

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

Proof. Suppose P is a point on a tangent to $\mathcal{Q}$, but not on $\mathcal{Q}$. Then the lines through P meet $\mathcal{Q}$ in at most two points, and thus they partition the points of $\mathcal{Q}$ into pairs and singletons. Each singleton determines a tangent to $\mathcal{Q}$ through P. Since $q+1$ is even, P lies on an even number of tangents. As P is on one tangent, it therefore lies on at least two. On the other hand, each pair of tangents to $\mathcal{Q}$ meet at a point off $\mathcal{Q}$, and this point is on two tangents. Thus there are at most $\binom{q+1}{2}$ triples formed from a pair of distinct

tangents and their point of intersection. This implies that any point off $\mathcal{Q}$ which is on a tangent is on exactly two. □

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

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

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

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

17.8 Segre's Characterisation of Conics

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

Let $\mathcal{C}$ be an oval in the projective plane over $GF(q)$. If $a \in \mathcal{C}$, let T_a be the linear function whose zero-set is the tangent to $\mathcal{C}$ at a. (Since there is a unique tangent at each point in the oval, T_a is well-defined.) Our proof depends on two lemmas.

17.8.1 Lemma. *If $\mathcal{C}$ is an oval in $PG(2,q)$ and q is odd, and a, b, c are distinct points on the oval*

$$T_a(b)T_b(c)T_c(a) = T_b(a)T_b(c)T_c(a)$$

Proof. We write the equation of the line in the plane through point u and v in the determinental form

$$[xuv] = 0.$$

We first consider the secants of $\mathcal{C}$ on c. Let w be a point on the oval distinct from a, b and c There are elements λ and μ of $\mathbb{F}$ such that

$$[xcw] = \lambda[xac] + \mu[xbc]$$

and substituting a and b for x yields

$$[acw] = \mu[abc], \quad [bcw] = \lambda[bac].$$

Consequently

$$[xcw] = \frac{[bcw]}{[bac]}[xac] + \frac{[acw]}{[abc]}[xbc],$$

from which we see that the line $c \vee w$ is determined by the ratio

$$\frac{[acw]}{[bcw]}.$$

Using the same techniques we have

$$T_c = \lambda[xac] + \mu[xbc]$$

where

$$T_c(a) = \mu[abc], \quad T_c(b) = \lambda[bac].$$

Hence

$$T_c = \frac{T_a(b)}{[bac]}[xac] + \frac{T_c(a)}{[abc]}[xbc]$$

and therefore the tangent at c is parameterized by the ratio $-T_c(a)/T_c(b)$. If $\mathcal{C}' := \mathcal{C} \setminus \{a, b, c\}$, the $q - 1$ terms in the product

$$-\frac{T_c(a)}{T_c(b)} \prod_{w \in \mathcal{C}'} \frac{[acw]}{[bcw]}$$

are exactly the non-zero elements of $\mathbb{F}$. Since q is odd it follows that this product is equal to -1 and consequently

$$T_c(a) \prod_{w \in \mathcal{C}'} [acw] = T_c(b) \prod_{w \in \mathcal{C}'} [bcw].$$

Similarly

$$T_a(b)\prod_{w\in\mathcal{C}'}[baw] = T_c(b)\prod_{w\in\mathcal{C}'}[caw]$$
$$T_b(c)\prod_{w\in\mathcal{C}'}[cbw] = T_c(b)\prod_{w\in\mathcal{C}'}[abw].$$

Multiplying our last three equations together, we get the result of the lemma. □

17.8.2 Lemma. *Let $\mathcal{C}$ be an oval in $PG(2,q)$, where q is odd. If a,b,c,d are four distinct points on $\mathcal{C}$, then*

$$T_b(a)T_c(b)T_a(d)[bcd] + T_c(b)T_a(b)T_b(d)[cad] + T_a(b)T_b(c)T_c(d)[abd] = 0.$$

Proof. We have

$$T_d = \lambda[xad] + \mu[xbd]$$

where

$$T_d(a) = \mu[abd], \quad T_d(b) = \lambda[bad]$$

and therefore

$$T_d = T_d(b)\frac{[xad]}{[bad]} + T_d(a)\frac{[xbd]}{[abd]}.$$

Straightforward rearrangements now yield

$$T_d(a)[bcd] + T_d(b)[cad] + T_d(c)[abd] = 0. \tag{17.8.1}$$

We now apply Lemma **??**. From that we find that

$$T_a(b)T_b(d)T_d(a) = T_b(a)T_d(b)T_a(d)$$

and therefore

$$\frac{T_a(b)T_b(d)}{T_b(a)T_d(b)}T_d(a) = T_a(d).$$

If we multiply (17.8.1) by $\frac{T_a(b)T_b(d)}{T_b(a)T_d(b)}$, we obtain

$$0 = T_a(d)[bcd] + \frac{T_a(b)}{T_b(a)}T_b(d)[cad] + \frac{T_a(b)T_b(d)T_d(c)}{T_b(a)T_d(b)}[abd];$$

since $T_c(b)T_b(d)T_d(c) = T_b(c)T_d(b)T_c(d)$, this yields

$$0 = T_a(d)[bcd] + \frac{T_a(b)}{T_b(a)}T_b(d)[cad] + \frac{T_a(b)T_b(c)}{T_b(a)T_c(b)}T_c(d)[abd].$$

The lemma follows at once. □

To prove Segre's theorem, view d as a variable in this lemma. Then $T_a(d)[bcd]$ is the equation for the singular conic consisting of the the tangent line to $\mathcal{C}$ at d and the line $b \vee c$. Similarly $T_b(d)[cad]$ is the equation for the conic consisting of the tangent at b and the secant $a \vee c$, and $T_c(d)[abd]$ is the equation for the conic consisting of the tangent at c and the secant $a \vee b$. It follows immediately that the equation is the statement of the lemma is a homogeneous quadratic and are done, almost. We must show it is not the zero polynomial.

For this, let e be the point of intersection of T_a and T_b. Since any point of an oval in odd characteristic is on 0 or 2 tangents, we have $T_c(e) \neq 0$. The tangent T_a meets $a \vee b$ in a, if e was on $a \vee b$ then T_a would have two points in common with $a \vee b$. Since this is impossible, we conclude that our quadratic does not vanish at e, and so cannot be the zero polynomial.

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

17.9 q-Arcs

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

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

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

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

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

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

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

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

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

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

Together these equations imply that

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

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

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

Exercises

1. Let $\mathcal{Q}$ be an oval in a projective plane of order q. Show that if q is odd then a k-arc can have at most $q+1$ points on it.

2. A k-arc in a projective plane *complete* if it is not a subset of a $(k+1)$-arc. Show that if there is a complete k-arc in a plane of order q then $q \leq \binom{k-1}{2}$.

3. Let $\mathbb{F}$ have even characteristic. Prove algebraically that the tangents to a smooth conic in $\mathbb{P}(2,\mathbb{F})$ are concurrent. (What can be said about the tangents to quadrics in $\mathbb{P}(n,\mathbb{F})$?)

4. Let p be a point in $\mathbb{P}(2,\mathbb{F})$ and let τ be a collineation in $PGL(3,\mathbb{F})$. Show that the points $l \cap l\tau$, as l ranges over the lines on p, lie on a conic. (This will probably not be easy.)

5. Show that an involution which fixes a smooth conic is a perspectivity. When is it an elation?

6. Let a, b and c form a 3-arc in $\mathbb{P}(2,\mathbb{F})$ and let l and m be tangents to this arc at a and b respectively. Show that there is a unique conic through these three points with l and m as tangents at a and b.

7. Let $\mathcal{D}(k)$ be the subset of the projective plane of the field $\mathbb{F}$ of order 2^n consisting of the points $(1, x, x^{2^k})^T$, where x ranges over the elements of $\mathbb{F}$, together with the point $(0,0,1)$. Show that this is an oval if $(k,n)=1$. (If k is not equal to 1 or $n-1$, this oval, together with its nucleus, does not coincide with any conic and its nucleus. For a proof of this, see Hirschfeld [PGOFF].)

Chapter 18

Polarities

In this chapter we study polarities of projective geometries.

18.1 Absolute Points

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

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

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

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

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

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

$$N^2 = (k-\lambda)I + \lambda J. \tag{18.1.1}$$

(Here J is the matrix with every entry equal to 1.) The number of absolute points of the polarity is equal to $\operatorname{tr} N$, which is in turn equal to the sum of the eigenvalues of N. From (18.1.1) we see that the eigenvalues of N^2 coincide with the eigenvalues of $(k-\lambda)I + \lambda J$. This means that N^2 must have as its eigenvalues

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

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

$$\operatorname{tr} N = k + (a-b)\sqrt{k-\lambda} \tag{18.1.2}$$

and, as $\operatorname{tr} N$, k, a and b are all integers, this implies that either $a = b$ or $(k-\lambda)$ is a perfect square. This proves (a) in the statement of the theorem. If the polarity is null then $\operatorname{tr} N = v$, whence (18.1.2) implies that

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

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

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

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

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

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

18.2 Polarities of Projective Planes

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

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

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

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

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

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

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

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

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

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

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

Assume now that n is even and let p be a non-absolute point. The $n + 1$ lines through p partition the remaining points of the plane. As each line must contain an absolute point ($n + 1$ is odd) there are at least $n + 1$ absolute points. Suppose that there are exactly $n + 1$ absolute points, and let x and y be two of them. If there is a non-absolute point q on $x \vee y$ then

the argument we have just shows that the n lines through q distinct from $x \vee y$ contain at least n distinct absolute points. Taken with x and y we thus obtain at least $n+2$ absolute points. This completes the proof of the theorem when n is even.

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

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

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

$$ns = \sum_{i=1}^{N} k_i. \tag{18.2.1}$$

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

$$s(s-1) = \sum_{i=1}^{N} k_i(k_i-1). \tag{18.2.2}$$

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

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

with equality if and only if the k_i are all equal. Using (**??**) and (18.2.2), this implies that $n^2 s \leq (s+n-1)N$. Recalling now that $N = n^2+n+1-s$ and indulging in some diligent rearranging, we deduce that $(s-1)^2 \leq n^3$, with equality holding if and only if the k_i are equal. This yields the theorem. □

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

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

We present a different proof of 18.2.3. Let N be the incidence matrix of a projective plane. We assume that our plane has a polarity and hence may assume that N is symmetric. From our calculations in 18.1 the eigenvalues of N are $q+1$ and $\pm\sqrt{q}$.

We can write N in partitioned form, with the absolute points and lines first:

$$N = \begin{pmatrix} I & M \\ M^T & A \end{pmatrix}$$

and note that this has quotient

$$B = \begin{pmatrix} 1 & q \\ x & q+1-x \end{pmatrix}$$

where, if there are m absolute points, $(q^2+q+1-m)x = mq$. The eigenvalues of the quotient are $q+1$ and $1-x = \mathrm{tr}(B) - q - 1$, and these must interlace the eigenvalues of N. Hence

$$-\sqrt{q} \le 1 - x = 1 - \frac{mq}{q^2+q+1-m}$$

Hence

$$m(q+\sqrt{q}+1) \le (q^2+q+1)(\sqrt{q}+1)$$

and accordingly

$$m \le (q-\sqrt{q}+1)(\sqrt{q}+1) = q^{3/2}+1.$$

If equality holds, the interlacing is tight and the partition is equitable, and $1-x = -\sqrt{q}$. Therefore each line contains either 1 or $\sqrt{q}+1$ absolute points. Each non-absolute point lies on exactly $\sqrt{q}+1$ absolute lines.

18.3 Polarities of Projective Spaces

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

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

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

$$A^{-1}(A^\tau)^T \mathbf{x}^{\tau^2} = \kappa \mathbf{x} \tag{18.3.1}$$

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

Since $A^{-1}(A^\tau)^T$ is a linear and not a semilinear mapping, it follows from (18.3.1) that $\mathbf{x}^{\tau^2}$ must lie in $V(n, \mathbb{F})$, and hence that $\tau^2 = 1$. Therefore (18.3.1) implies that $A^{-1}(A^\tau)^T = \kappa I$ and so we have shown that every polarity is determined by a field automorphism τ of order dividing two and a linear mapping A such that $(A^\tau)^T = \kappa A$. Now

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

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

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

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

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

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

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

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

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

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

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

If ϕ is a polarity on the vector space V, then the map

$$(x,y) \mapsto x^\phi(y)$$

is a bilinear form. A subspace U of V is non-singular if $U \cap U^\perp = 0$. In this case $U^\perp$ is a complement to U in V and $U + U^\perp = V$.

18.4 Polar Spaces

A *polar space* is an incidence structure such that:

(a) If p is a point and ℓ is a line not on p, either there is a unique point on ℓ collinear with p or all points on ℓ are collinear with p.

(b) Each line is incident with at least three points.

If we only require that each line contains at least two points we have a *generalized polar space*. The polar space is *nondegenerate* if there is no point that is collinear with all the other points. A polar space is not defined

to be a partial linear space, but if it is nondegenerate then it does follow from the axioms that it is a partial linear space.

We view a point as collinear with itself. If x is a point then $x^\perp$ is the set of all points collinear with x; if S is a set of points then

$$S^\perp = \bigcap_{x \in S} x^\perp.$$

A set of points S is a subspace of a polar space if each pair of points in it are collinear and each line that contains two points of S is itself contained in S. Note that, for subspaces of partial linear spaces, we did not require that each pair of points be collinear. The *rank* of a subspace is its height in the poset of subspaces. The rank of a polar space is the maximum rank of a subspace.

18.5 Symplectic Spaces

We start with two very useful results, valid for all forms.

18.5.1 Lemma. *Let $\langle \cdot, \cdot \rangle$ be a non-degenerate form on V. If U is subspace of V and $U \cap U^\perp = 0$, the restriction of the form to U is non-degenerate.*

Proof. Exercise. □

A subspace U is isotropic relative to a form if the restriction of the form to U is the zero form

18.5.2 Lemma. *If U is an isotropic subspace of dimension k in a space of dimension d, then $2k \leq d$.*

Proof. If $\dim(U) = k$ then $\dim(U^\perp) = d - k$ and if $U \leq U^\perp$ we must have $k \leq d - k$. □

If the assumption of the lemma holds then the restriction of the form to $U^\perp$ is also non-degenerate, and we have useful direct sum decomposition $V = U \oplus U^\perp$. In fact we can view the form as a sum of two forms, one on U and the other on $U^\perp$.

Let V be a vector space of dimension n over $GF(q)$ and let H be a matrix over $GF(q)$ such that $H^T = -H$ and, if the characteristic of our field is two, than all diagonal entries are zero. Then

$$(x, y) \mapsto x^T H y$$

is a symplectic form on V; it is non degenerate if and only if H is invertible. A standard example is

$$H = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}.$$

Define $x^\perp$ to be the kernel of $x^T H$. Since

$$x^T H x = -x^T H^T x = -x^T H x$$

we see that $x^T H x = 0$ for all x, and all 1-dimensional subspaces are isotropic.

If H is non-singular then $x^\perp$ has codimension one, and so there is a vector y in V such that $y^T H x \neq 0$. The subspace spanned by x and y is non-singular, and hence V can be written as a direct sum $U \oplus U^\perp$, where U is the subspace spanned by x and y. Since U is non-singular, it follows that $U^\perp$ is non-singular and we deduce by induction on the dimension that there are vectors

$$x_1, y_1, \ldots, x_m, y_m$$

such that $x_i^T H y_i = 1$ for all i and, if $i \neq j$ then

$$x_i^T H x_j = x_i^T H y_j = y_i^T H y_i = 0.$$

This implies that $n = 2m$. And also that two non-degenerate symplectic spaces of the same dimension over $GF(q)$ are isomorphic.

18.6 Symplectic Spreads

A non-degenerate symplectic space with dimension $n = 2m$ contains spreads. A *symplectic spread* is a spread whose components are isotropic spaces. Suppose our symplectic form is given by

$$H := \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

Then the column spaces of the matrices

$$\begin{pmatrix} I \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ I \end{pmatrix}$$

are both isotropic. Now

$$\begin{pmatrix} I & A^T \end{pmatrix} H \begin{pmatrix} I \\ A \end{pmatrix} = A - A^T$$

and so the column space of

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

is isotropic relative to H if and only if A is symmetric.

18.7 Unitals

Let V be a vector space over $\mathbb{F}$ and let σ be an automorphism of $\mathbb{F}$ with order two. A form $\langle x, y\rangle$ on V is *unitary* relative to σ if it is linear in the second variable and

$$\langle y, x\rangle = \langle x, y\rangle^\sigma.$$

Note that the form is semilinear in the first variable:

$$\langle cx, y\rangle = c^\sigma \langle x, y\rangle.$$

If $x \in V$ then $x^\perp$ is defined by

$$x^\perp := \{y : \langle x, y\rangle = 0\}.$$

Similarly we define $U^\perp$ for a subspace U of V If $a^\perp = V$ if and only if $a = 0$, we say that the form is non-degenerate. The happens if and only if A is invertible. If the form is non-degenerate, then $(U^\perp)^\perp = U$ and $\dim(U^\perp) = \dim(V) - \dim(U)$.

We give a construction. An $n \times n$ matrix A over $\mathbb{F}$ is σ-Hermitian if

$$(A^\sigma)^T = A.$$

Thus the identity matrix is σ-Hermitian. We define a form on the vector space $\mathbb{F}^n$ by

$$\langle x, y\rangle := (x^\sigma)^T A y.$$

Then

$$\langle x, y\rangle^\sigma = x^T A^\sigma y^\sigma = (y^\sigma)^T (A^\sigma)^T x$$

and therefore if A is σ-Hermitian, then

$$\langle x, y\rangle^\sigma = \langle y, x\rangle.$$

This form is non-degenerate if and only if A is invertible. It is semilinear in the first variable and linear in the second. If A is σ-Hermitian and invertible, we call $\langle x, y\rangle$ a σ-Hermitian form on V. (And before long we will drop the reference to σ.)

The set of vectors x such that $\langle x, x\rangle = 0$ is called a *Hermitian variety*.

We are only concerned with finite fields, and in this case any field automorphism of order two arises as the q-th power map on a field of order q^2. We may take $A = I$, and then

$$\langle x, y\rangle = \sum_i x_i^q y_i.$$

If $\mathbb{F}$ is our field and $\mathbb{F}_0$ is its subfield of order q, then the map $x \mapsto x^{q+1}$ is the norm relative to $\mathbb{F}_0$; it is a surjective homomorphism from $\mathbb{F}^*$ onto $\mathbb{F}_0^*$. We denote the norm of x by $N(x)$ or, if more precision is needed, by $N_{\mathbb{F}/\mathbb{F}_0}(x)$.

18.8 Generalized Quadrangles

Let V be a vector space of dimension four over the $\mathbb{F}$ of order q, where we view the elements of V as pairs (u, v) of 2-dimensional vectors. If we set

$$\langle (a, b), (c, d)\rangle := a^T d - b^T c$$

we have a non-degenerate alternating form on V, which corresponds to the matrix

$$H = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

The points an 2-dimensional isotropic subspaces of $\mathcal{P}(V)$ form a GQ which we will denote by $W(q)$. We determine its parameters. Since a maximal isotropic subspace has dimension two over $\mathbb{F}$, it contains $q+1$ 1-dimensional subspaces, whence $s = q$. If $u \in V$ the $\dim(u^\perp) = 3$ and each 2-dimensional subspace of $u^\perp$ that contains x is an isotropic subspace that contains x. It follows that there are exactly $q + 1$ 2-dimensional isotropic subspaces that contain u, and therefore $t = q$.

For a second example, take the isotropic points relative to a non-degenerate unitary form on $PG(3, q^2)$. Denote this set of points by $\mathcal{H}$. Since the form is non-degenerate, there is no plane contained in $\mathcal{H}$. If $u \in \mathcal{H}$ then the restriction of the form to $u^\perp$ is degenerate and hence $u^\perp$ contains an isotropic line which must contain u. It follows that the isotropic points and lines form a $GQ(q^2, q)$ and therefore there are $(q^2+1)(q^3+1)$ points and $(q+1)(q^3+1)$ lines.

18.9 Uniqueness

We show that unitary forms are unique, up to a change of basis.

18.9.1 Theorem. *If γ is a non-degenerate unitary form on a vector space V over a finite field, then V has an orthonormal basis.*

Proof. We first show that there is a non-isotropic vector x_1. Assume by way of contradiction that $\gamma(x, x) = 0$ for all x in V. Then for all x and y

$$0 = \gamma(x+y, x+y) = \gamma(x, x) + \gamma(y, y) + \gamma(x, y) + \gamma(y, x) = \gamma(x, y) + \gamma(y, x).$$

If $a \in \mathbb{F}$, then

$$0 = \gamma(x, ay) + \gamma(ay, x) = a\gamma(x, y) + a^\sigma \gamma(y, x)$$

and hence

$$(a^\sigma - a)\gamma(x, y) = 0$$

for all a in $\mathbb{F}$. Therefore $\gamma(x, y) = 0$ and γ is the zero form.

Thus there is non-isotropic vector x_1 in V. Let U denote the span of x_1. Then the restriction of γ to U is non-degenerate, as is its restriction to $U^\perp$. By induction on its dimension, $U^\perp$ has an orthogonal basis, and the union of this basis with x_1 is orthogonal.

We convert our orthogonal basis to an orthonormal basis. We have

$$\gamma(ax, ax) = aa^\sigma \gamma(x, x) = a^{q+1}\gamma(x, x).$$

Since $\gamma(x, x) \in \mathbb{F}_0$ and since the norm map is onto, if $x \neq 0$ we can choose a so $\gamma(x, x) = 1$. Therefore V has an orthonormal basis. □

It follows that if $\mathbb{F}$ is finite, then we are free to assume that our unitary form is given by

$$\langle x, y \rangle = \sum_i x_i^{1+q} y_i.$$

18.10 Hermitian Geometry in the Plane

We are interested in the geometry of the isotropic points of a unitary polarity on a vector space over a finite field.

Although projective lines are simple, we need to determine what happens with them. Suppose $\dim(V) = 2$ and γ is a unitary polarity on V. Assume $\{x, y\}$ is an orthonormal basis for V. Then

$$\gamma(x + ty, x + ty) = 1 + t^{q+1}$$

and hence there are $q+1$ elements t (in our field of order q^2) such that $x+ty$ is isotropic. Suppose there is no orthonormal basis. Then γ is degenerate and there is a point y such that $\gamma(y, x) = 0$ for all x. If γ is not the zero form, then there is a non-isotropic point x such that $\gamma(x, y) = 0$ and consequently

$$\gamma(tx + y, tx + y) = t^{q+1}\gamma(x, x) + t^q\gamma(x, y) + t\gamma(y, x) = t^{q+1}\gamma(x, x) \neq 0.$$

Hence y is the only isotropic point on ℓ (and $\ell \subseteq y^\perp$). Thus we have:

18.10.1 Lemma. *If ℓ is a line in a projective space over a field of order q^2 with a unitary polarity, then the number of isotropic points on ℓ is 1, $q + 1$ or $q^2 + 1$.* □

18.10.2 Lemma. *If ℓ is a line that contains exactly one isotropic point (relative to a unitary form on a projective space), then the isotropic point is $\ell^\perp$.*

Proof. Suppose $x \in \ell$ and $\langle x, x\rangle = 0$. If $\ell^\perp \in \ell$ and $\ell^\perp \neq x$ then ℓ and x span ℓ and ℓ is isotropic. □

We turn next to projective planes.

18.10.3 Theorem. *If γ is a non-degenerate unitary polarity on the projective plane over a field of order q^2, then it has exactly $q^3 + 1$ isotropic points.*

Proof. The first step is to show that there are at least two isotropic points. By Lemma **??** there is an isotropic point x. Suppose ℓ is a line on x not equal to $x^\perp$. Then ℓ is not absolute and so contains exactly $q + 1$ absolute points. Therefore the total number of absolute points is $1 + q^3$. □

The isotropic points relative to a non-zero degenerate unitary polarity form either a single line, or $q + 1$ collinear lines.

18.11 Quadratic Spaces and Polarities

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

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

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

Let β be the bilinear form defined by

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

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

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

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

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

We are going to classify quadratic forms over finite fields. For any subspace W of V, we define

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

18.12 Perspectivities of Polar Spaces

Suppose we have a sesquilinear form $\langle \cdot, \cdot \rangle$ on the vector space V, that is, a form semilinear in its first coordinate and linear in its second. If

$$a^\perp := \{x : \langle a, x \rangle = 0\}$$

then the mapping $a \mapsto a^\perp$ is a polarity on the projective space $\mathcal{P}(V)$ (and all polarities arise in this way). Given the form, we define a map τ_a from V to itself by

$$\tau_a : x \mapsto x + \lambda \langle a, x \rangle a.$$

This is linear (in x) because it is the sum of two linear mappings and it fixes each vector in $a^\perp$. It is invertible if $\lambda \langle a, a \rangle \neq -1$; in this case we see that it is a perspectivity.

We are interested in seeing when this perspectivity is compatible with the polarity. So if $y \in x^\perp$, we want to know if $\tau_a(x) \in \tau_a(y)^\perp$. This will certainly hold if

$$\langle \tau_a(x), \tau_a(y) \rangle = \langle x, y \rangle$$

for all x and y.

There are two cases. Suppose first that $\langle a, a \rangle = 0$. Then

$$\langle \tau_a(x), \tau_a(y) \rangle - \langle x, y \rangle = \overline{\lambda \langle a, x \rangle} \langle a, y \rangle + \lambda \langle x, a \rangle \langle a, y \rangle$$

and this is zero if our form is symplectic, or if it is unitary and $\overline{\lambda} = -\lambda$. You should prove that τ_a fixes any hyperplane that contains a. (Hence τ_a is an elation.)

Otherwise, assume that our form is associated to a quadratic form Q. If $Q(a) \neq 0$ and we set $\lambda = Q(a)^{-1}$, then

$$Q(\tau_a(x)) = Q(x) + \lambda^2 \langle a, x \rangle^2 Q(a) + \lambda \langle a, x \rangle \langle x, a \rangle.$$

Since the form is symmetric it follows that $Q(\tau_a(x)) = Q(x)$; since τ_a preserves the quadratic form, it must also preserve the associated bilinear form. You should prove that $\tau_a^2 = 1$ and $\tau_a(a) = -a$.

Chapter 19

Polar Spaces

Chapter 20

Reguli, Lines and Spreads

It might appear that there is very little to said about lines. They do seem straightforward objects. Our starting point here will be the sets of $q+1$ lines which meet a given a set of three pairwise skew lines in $PG(3,q)$.

20.1 Reguli

We saw that every set of three pairwise skew lines in $\mathbb{P}(3,\mathbb{F})$ lies in a unique hyperbolic quadric. This quadric contains altogether $2(q+1)$ lines—there are $q+1$ lines which meet each of the first three lines, and if we take three of these $q+1$ then they are each met in three points by another set of $q+1$ lines, including the three we started with. A *regulus* in $\mathbb{P}(3,\mathbb{F})$ is a set of $q+1$ pairwise skew lines lying on a hyperbolic quadric. We can equivalently identify the regulus with the corresponding set of $q+1$ pairwise skew 2-dimensional subspaces of $V(4,\mathbb{F})$. (The definition of a regulus can be extended to spaces of higher dimension; however the one just given will suffice for now.) Each hyperbolic quadric determines two reguli. We say that these reguli are *opposite*. Opposite reguli cover the same set of $(q+1)^2$ points, but have no lines in common.

From our work on hyperbolic quadrics in $\mathbb{P}(3,\mathbb{F})$, we know that every set of three pairwise skew lines lies in a unique regulus. If a spread contains a regulus $\mathcal{R}$ then we may replace it by its opposite regulus $\mathcal{R}'$. The result is a new spread. This idea can be used to construct new translation planes, as we shall see. A spread is *regular* if, whenever it contains three lines l_1, l_2 and l_3 from a regulus, it contains all the lines in it. Over $GF(2)$, every set

of three pairwise skew lines is a regulus, and hence all spreads are regular. In this case, regularity will be of no use to us.

20.1.1 Lemma. *Let U be a vector space of dimension two over $\mathbb{F}$ and suppose $V = U \oplus U$. Let σ_i for $i = 1, 2, 3, 4$ be distinct elements of $GL(U)$. Then the subspaces $V(\sigma_i)$ lie on a regulus if and only if*

$$(\sigma_4 - \sigma_2)^{-1}(\sigma_4 - \sigma_1)(\sigma_3 - \sigma_1)^{-1}(\sigma_3 - \sigma_2) = \kappa I$$

for some non-zero element κ of $\mathbb{F}$.

Proof. Let α be the element

$$\begin{pmatrix} \sigma_2 & -\sigma_1 \\ -1 & 1 \end{pmatrix}$$

of $GL(V)$. Then α maps $V(\sigma_1)$ onto $V(0)$ and $V(\sigma_2)$ onto $V(\infty)$. The subspaces $V(\sigma_3)$ and $V(\sigma_4)$ are mapped respectively to $V(\sigma_2 - \sigma_3)^{-1}(\sigma_3 - \sigma_1))$ and $V(\sigma_2 - \sigma_4)^{-1}(\sigma_4 - \sigma_1))$. Thus we need a condition for these two subspaces to lie on a regulus with $V(0)$ and $V(\infty)$. We claim that $V(\rho)$ is on a regulus with $V(0)$, $V(\infty)$ and $V(\tau)$ if and only if $\rho = \kappa\tau$ for some non-zero element κ of $\mathbb{F}$. We work in the projective space determined by V. Suppose that $a \in V(\rho)$. Then there is a unique line through a meeting both $V(0)$ and $V(\infty)$; since our four spaces are part of a regulus this line must meet $V(\tau)$ in some point b. We may assume b is represented by the vector $(u, u\tau)$, where $u \in U$, and we observe that

$$(u, u\tau) = (u, 0) + (0, u\tau).$$

Thus $a \vee b$ must meet $V(0)$ in the point represented by $(u, 0)$ and $V(\infty)$ in the point belonging to $(0, u\rho)$. If a is represented by $(v, v\rho)$ for some v in U, it follows that

$$(v, v\rho) = \lambda_u(u, 0) + \mu_u(0, u\tau). \tag{20.1.1}$$

Here λ_u and μ_u are elements of $\mathbb{F}$, and they cannot be zero since a is not in $V(0)$ or $V(\infty)$. Hence (20.1.1) implies that $v = \lambda_v u$ and $u\rho = (\mu_u/\lambda_u)u\tau$. Assume $\kappa(u) = \mu_u/\lambda_u$. Then, for all u in U,

$$u\rho\tau^{-1} = \kappa(u)u.$$

This implies that every vector in U is an eigenvector of $\tau\rho^{-1}$ and hence that the matrix must have the form κI for some non-zero scalar κ. Thus $\tau = \kappa\rho$

if $V(\rho)$ and $V(\tau)$ lie on a regulus with $V(0)$ and $V(\infty)$. It is routine to verify conversely that if $\tau = \kappa\rho$ then $V(\rho)$ and $V(\tau)$ lie on a regulus with $V(0)$ and $V(\infty)$. Therefore $V(\sigma_2 - \sigma_3)^{-1}(\sigma_3 - \sigma_1))$ and $V(\sigma_2 - \sigma_4)^{-1}(\sigma_4 - \sigma_1))$ both lie on a regulus with $V(0)$ and $V(\infty)$ if and only if, for some non-zero element κ of $\mathbb{F}$,

$$(\sigma_2 - \sigma_4)^{-1}(\sigma_4 - \sigma_1) = \kappa(\sigma_2 - \sigma_3)^{-1}(\sigma_3 - \sigma_1).$$

On rearranging this, we obtain the desired result. □

It follows from the above proof that the regulus containing $V(0)$, $V(\infty)$ and $V(\sigma)$ consists of all the lines $V(\lambda\sigma)$, where $\lambda \in \mathbb{F}$.

20.1.2 Theorem. *If $|\mathbb{F}| > 2$ then a spread in $\mathbb{P}(3, \mathbb{F})$ is regular if and only if the plane it determines is Pappian.*

Proof. Suppose that Σ determines a regular spread in $\mathbb{P}(3, \mathbb{F})$. We may assume without loss that $I \in \Sigma$. If $\sigma \in \Sigma$ then all lines in the regulus containing $V(0)$, $V(\infty)$ and $V(\sigma)$ must belong to the spread. From the proof of the previous lemma we see that $\kappa\sigma \in \Sigma$ for all non-zero elements κ of $\mathbb{F}$. All lines in the regulus containing $V(\infty)$, $V(\rho)$ and $V(\sigma)$ must belong to the spread. Suppose that $V(\tau)$ is one of these lines. The linear mapping

$$\alpha = \begin{pmatrix} \rho & -I \\ -I & 0 \end{pmatrix}$$

sends $V(\infty)$, $V(\rho)$, $V(\sigma)$ and $V(\tau)$ to $V(0)$, $V(\infty)$, $V((\sigma-\rho)^{-1})$ and $V((\tau-\rho)^{-1})$ respectively. Since α must map reguli to reguli, we deduce that

$$(\sigma - \rho)^{-1} = \kappa(\tau - \rho)^{-1}$$

for some non-zero element κ of $\mathbb{F}$. Taking the inverse of each side and rearranging yields

$$\tau = \kappa\sigma - (\kappa - 1)\rho.$$

We know already that Σ is closed under multiplication by non-zero elements of $\mathbb{F}$; with this last equation we now deduce that $\Sigma \cup 0$ is closed under addition. Hence it is a vector space over $\mathbb{F}$. In fact it is a 2-dimensional vector space over $\mathbb{F}$.

If $\mathbb{F}$ is finite then a simple cardinality argument shows that $\Sigma \cup 0$ must have dimension two over $\mathbb{F}$. More generally, we proceed as follows. The 2×2

matrices over $\mathbb{F}$ form a vector space of dimension four. The matrices with determinant zero form a hyperbolic quadric in the corresponding projective space. (The proof of this is left as an exercise.) A hyperbolic quadric in $\mathbb{P}(3,\mathbb{F})$ contains lines and, since every line in $\mathbb{P}(3,\mathbb{F})$ meets every hyperplane, it follows that a subspace formed by non-singular matrices must have rank at most two. Hence there is $\sigma \in \Sigma$ such that σ and I form a basis for Σ over $\mathbb{F}$. Thus every element of Σ is a linear combination of σ and I, and so multiplication of elements of Σ is commutative. By Lemma 3.2.5 we now deduce that $\mathbb{F}$ is Pappian. If $|\mathbb{F}| = 2$ then it is easy to show that all spreads in $V(4,2)$ give Desarguesian planes. (One possibility is to note that there is only one projective plane of order four.) The proof that a spread in $\mathbb{P}(3,\mathbb{F})$ is regular if the plane it determines is Pappian is left as an exercise. □

In the proof of this theorem we made use of the techniques developed in proving ??Lemma 1.1. If we are prepared to accept assertions of the form

$$(\tau - \infty)^{-1}(\sigma - \infty) = 1$$

then we could have appealed to the result of ??Lemma 1.1 instead. This would have been less work, since we would not have needed to use the linear mapping α.

??Theorem 1.2 has an important consequence, which we develop in two steps.

20.1.3 Corollary. *Let $\mathcal{S}$ and $\mathcal{S}'$ be two distinct spreads in $V(4,q)$. If $\mathcal{S}$ and $\mathcal{S}'$ have four components in common, not all lying on a single regulus, then the spreads are not both regular.*

Proof. We may assume that both spreads contain $V(0)$, $V(1)$, $V(\infty)$ and $V(\sigma)$. But every component of a regular spread containing these components is of the form $V(\tau)$, where τ is a linear combination of σ and I. Thus the two spreads cannot be both regular and distinct. □

Now take a regular spread $\mathcal{S}$ in $V(4,\mathbb{F})$. This spread contains a regulus; delete the lines in it and replace them with the lines of the opposite regulus. The result is a spread $\mathcal{S}'$ having $q^2 - q$ components in common with $\mathcal{S}$. As $q^2 - q > q + 1$ the previous corollary implies that the new spread is not regular. Thus we have shown that there is a non-Desarguesian translation plane of order q^2 for all prime powers q. It should be clear that we could simultaneously replace several disjoint reguli by their opposites. If 'several'

is t, the new spreads would not be Desarguesian provided $q^2 + 1 - t(q + 1) > q + 1$. The problem that remains is to decide when these spreads are isomorphic.

20.2 Plücker Coordinates of Lines

Any line in $\mathbb{P}(3, \mathbb{F})$ can always be represented by a pair of vectors, corresponding to two distinct points on it. One difficulty with this is that the same line can be represented in many different ways. An unambiguous representation is available. Suppose that $\ell = a \vee b$. Consider the matrix

$$\begin{pmatrix} a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 \end{pmatrix}$$

with rows representing a and b. For distinct i and j, define

$$\ell_{ij} = a_i b_j - a_j b_i.$$

The numbers ℓ_{ij} are the *Plücker coordinates* of ℓ. If c and d are a second set of distinct points on ℓ then we find that

$$\begin{pmatrix} c_0 & c_1 & c_2 & c_3 \\ d_0 & d_1 & d_2 & d_3 \end{pmatrix} = M \begin{pmatrix} a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 \end{pmatrix}$$

for some 2×2 matrix M. The Plücker coordinates for ℓ computed using c and d will thus be equal to $\det(M)$ times the corresponding coordinates ℓ_{ij}. Note also that

$$ab^T - ba^T = \begin{pmatrix} 0 & \ell_{01} & \ell_{02} & \ell_{03} \\ -\ell_{01} & 0 & \ell_{12} & \ell_{13} \\ -\ell_{02} & -\ell_{12} & 0 & \ell_{23} \\ -\ell_{03} & -\ell_{13} & -\ell_{23} & 0 \end{pmatrix}.$$

This is a skew symmetric matrix whose column space is the span of a and b; its kernel consists of all vectors x such that $a^T x = b^T = 0$, in other words it is $\ell^\perp$. We denote this matrix by $S(\ell)$. Since a skew symmetric matrix has even rank we see that $\mathrm{rk}(S(\ell)) = 2$ for any line.

If we define

$$\overline{S}(\ell) := \begin{pmatrix} 0 & -\ell_{23} & \ell_{13} & -\ell_{12} \\ \ell_{23} & 0 & -\ell_{03} & \ell_{02} \\ -\ell_{13} & \ell_{03} & 0 & -\ell_{01} \\ \ell_{12} & -\ell_{02} & \ell_{01} & 0 \end{pmatrix}$$

then

$$S(\ell)\overline{S}(\ell) = (\ell_{01}\ell_{23} - \ell_{02}\ell_{13} + \ell_{03}\ell_{12})I$$

and thus we deduce that the Plücker coordinates of a line satisfy

$$\ell_{01}\ell_{23} - \ell_{02}\ell_{13} + \ell_{03}\ell_{12} = 0. \tag{20.2.1}$$

Conversely, if this condition holds then $\mathrm{rk}(S(\ell)) = 2$. Thus we have a bijection between the lines of $PG(3,\mathbb{F})$ and the 4×4 skew symmetric matrices over $\mathbb{F}$ with rank two.

Equivalently. we have shown that the numbers ℓ_{ij} are the Plücker coordinates of a line if and only if (20.2.1) holds, and that they determine the line. (Which is why they are called coordinates.)

You might show that

$$\det(S(\ell)) = \det(\overline{S}(\ell)) = (\ell_{01}\ell_{23} - \ell_{02}\ell_{13} + \ell_{03}\ell_{12})^2.$$

20.3 The Klein Quadric

There is another important observation to be made. The set of points in $\mathbb{P}(5,\mathbb{F})$ satisfying

$$x_0x_5 - x_1x_4 + x_2x_3 = 0$$

is a smooth quadric containing the subspace $x_0 = x_1 = x_2 = 0$. Thus it has index three, and is therefore hyperbolic. It is called the *Klein quadric*. We have shown that the lines in $\mathbb{P}(3,\mathbb{F})$ correspond bijectively to the points on the Klein quadric in $\mathbb{P}(3,\mathbb{F})$. We wish to investigate the relation between the geometry of this quadric and the lines in $\mathbb{P}(3,\mathbb{F})$. If ℓ is a line in $\mathbb{P}(3,\mathbb{F})$, let $\hat{\ell}$ denote the corresponding point on the Klein quadric. Denote the Klein quadric by $\mathcal{K}$.

Our first problem is to decide which sets of lines in $\mathbb{P}(3,\mathbb{F})$ correspond to the lines contained in the Klein quadric. The lines $a \vee b$ and $c \vee d$ are skew if and only if

$$\det\begin{pmatrix} a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 \\ c_0 & c_1 & c_2 & c_3 \\ d_0 & d_1 & d_2 & d_3 \end{pmatrix} \neq 0 \tag{20.3.1}$$

Denote the Plücker coordinates of $a \vee b$ and $c \vee d$ respectively by ℓ_{ij} and m_{ij}. Expanding the determinant in (20.3.1), we obtain

$$\ell_{01}m_{23} - \ell_{02}m_{13} + \ell_{03}m_{12} + \ell_{12}m_{03} - \ell_{13}m_{02} + \ell_{23}m_{01} = 0.$$

If we view $a \vee b$ as a fixed line and $c \vee d$ as varying, this shows that the lines which meet $a \vee b$ in $\mathbb{P}(3, \mathbb{F})$ correspond to the points on the intersection of the Klein quadric with the hyperplane with coordinate vector

$$(\ell_{23}, -\ell_{13}, \ell_{12}, \ell_{03}, -\ell_{02}, \ell_{01}).$$

This hyperplane contains the point corresponding to $a \vee b$, and a straightforward calculation shows that it is the tangent hyperplane to the quadric at this point. Thus ℓ meets m if and only if $\hat{\ell} \vee \hat{m}$ is contained in $\mathcal{K}$. It follows that the set of lines passing through a fixed point in $\mathbb{P}(3, \mathbb{F})$ determine a subspace of rank three in $\mathcal{K}$.

A second class of rank three subspaces is provided by the sets of lines which lie in a given plane in $\mathbb{P}(3, \mathbb{F})$. There are $[4] = q^3+q^2+q+1$ subspaces of each type. Since we already know that a hyperbolic quadric in $\mathbb{P}(5, \mathbb{F})$ contains

$$2(1 + q)(1 + q^2)$$

subspaces of rank three, we have therefore found all rank three subspaces on $\mathcal{K}$. Note that any two subspaces of the same type have a unique point in common, while two subspaces of different types are either disjoint or meet in a line. Any line in $\mathcal{K}$ is contained in a rank three subspace on $\mathcal{K}$. Hence a line in $\mathcal{K}$ corresponds to $q + 1$ concurrent coplanar lines in $\mathbb{P}(3, \mathbb{F})$. If ℓ and m are two skew lines in $\mathbb{P}(3, \mathbb{F})$ then $\hat{\ell}$ and $\hat{m}$ are not collinear in $\mathcal{K}$ and $T_{\hat{\ell}} \cap T_{\hat{m}}$ is a quadric in a subspace of rank four with $(q + 1)^2$ points. By ??Corollary 6.1.2 and Lemma 6.3.1 it is smooth, and thus must be a hyperbolic quadric. If n is a third line in $\mathbb{P}(3, \mathbb{F})$ skew to ℓ and m then the intersection of $T_{\hat{\ell}}$, $T_{\hat{m}}$ and $T_{\hat{n}}$ is a conic. (Since $\hat{n}$ is not collinear in $\mathcal{K}$ with $\hat{\ell}$ or $\hat{m}$, the intersection cannot contain a line.) Thus each regulus in $\mathbb{P}(3, \mathbb{F})$ corresponds to a plane in $\mathbb{P}(5, \mathbb{F})$ meeting $\mathcal{K}$ in a smooth conic. Each spread of $\mathbb{P}(3, \mathbb{F})$ determines a set of $q^2 + 1$ points on $\mathcal{K}$. This set cannot contain two points lying on a line of $\mathcal{K}$, therefore no two of its points lie on a line of $\mathcal{K}$, and no three of its points are collinear in $\mathbb{P}(5, \mathbb{F})$. Each line of $\mathbb{P}(3, \mathbb{F})$ lies on $q + 1$ points and in $q + 1$ planes of $\mathbb{P}(3, \mathbb{F})$; hence each point of $\mathcal{K}$ lies on $2(q + 1)$ maximal subspaces. Therefore any set of points of $\mathcal{K}$ which meets every subspace of rank three must have at least

$$2(q + 1)(q^2 + 1)/2(q + 1) = q^2 + 1$$

points in it. If $\mathcal{O}$ is a set of $q^2 + 1$ points meeting each maximal subspace then there must be exactly one point of it in each maximal subspace. Hence

no two points of $\mathcal{O}$ can lie on a line of $\mathcal{K}$, and so no three points of $\mathcal{O}$ can be collinear in $\mathbb{P}(5,\mathbb{F})$. Thus it determines a spread in $\mathbb{P}(3,\mathbb{F})$. An *ovoid* in a quadric is a set of points which has exactly one point in each maximal subspace. We have just proved that an ovoid in the Klein quadric has exactly q^2+1 points, and corresponds to a spread in $\mathbb{P}(3,\mathbb{F})$.

20.3.1 Lemma. *A spread in $\mathbb{P}(3,\mathbb{F})$ determines a Desarguesian plane if and only if the ovoid it determines in the Klein quadric is contained in a subspace of $\mathbb{P}(5,\mathbb{F})$ with rank four.*

Proof. The case $|\mathbb{F}| = 2$ is left as an exercise; henceforth we assume that $|\mathbb{F}| > 2$. Let $\mathcal{S}$ be a spread in $\mathbb{P}(3,\mathbb{F})$ and suppose that $\hat{\mathcal{S}}$ is contained in a subspace H of rank four in $\mathbb{P}(5,\mathbb{F})$. Then $H \cap \mathcal{K}$ is a quadric in $\mathbb{P}(3,\mathbb{F})$ and, as it contains an ovoid, it can only be an elliptic quadric. As $|\mathcal{S}| = q^2+1$, it follows that $\hat{\mathcal{S}}$ must be an elliptic quadric. Since any plane section of an elliptic quadric in $\mathbb{P}(3,\mathbb{F})$ is either a point or a smooth conic, we deduce that $\mathcal{S}$ must be regular.

We now assume conversely that $\mathcal{S}$ is a regular spread, and seek to show that its image in $\mathcal{K}$ spans a space of rank four. We may assume without loss that the spread contains the subspaces $V(0)$, $V(\infty)$, $V(I)$ and $V(\rho)$, for some matrix ρ. If $V(\sigma)$ is another component of the spread then the proof of Theorem 1.2 shows that for some scalar κ, the subspace $V(\kappa\sigma)$ lies in a regulus with $V(\infty)$, $V(I)$ and $V(\tau)$. Similarly, $V(\sigma)$ lies in a regulus with $V(0)$, $V(\infty)$ and $V(\kappa\sigma)$. The image of any regulus in $\mathbb{P}(3,\mathbb{F})$ is contained in a plane in $\mathbb{P}(5,\mathbb{F})$, hence we deduce that the image of $V(\sigma)$ in $\mathcal{K}$ lies in the span of the images of $V(0)$, $V(I)$, $V(\infty)$ and $V(\rho)$. Therefore the image of our spread lies in a subspace of rank four. □

20.4 Reguli in Higher Dimensions

Our definition of a regulus can be extended to vector spaces of even dimension greater than four. Consider the pairwise skew subspaces U, V and W of rank m in $\mathbb{P}(2m-1,\mathbb{F})$. If w is a point not in $U \cup V$ then $w \vee V$ is a subspace of rank $m+1$ and hence must meet U in a point. The line joining this point to w must meet V, because V is a hyperplane in $w \vee V$. Thus, for each point w not in $U \cap V$, there is a unique line through it which meets both U and V. If we now allow w to range over the points of W, we thus

obtain a set of pairwise skew lines, each meeting U, V and W in exactly one point. We will define a *regulus* in $V(2m, \mathbb{F})$ to be a set of pairwise skew subspaces of dimension m, indexed by the elements of $\mathbb{F}$, such that any line which meets three of these subspaces is contained in their union. Regular spreads are defined as before. Over the field with two elements they are still uninteresting. Now suppose that $V = V(2m, \mathbb{F})$ and that ρ is an element of $GL(V)$ such that U and $U\rho$ are skew. If $u \in U$ let $L(u)$ be the line $u \vee u\rho$ and if $\kappa \in \mathbb{F}$, let $U(\kappa)$ be the subspace $\{u + \kappa u\rho : u \in U\}$. (Here $U(\infty) = U\rho$.) It is easy to see that

$$\bigcup_{u \in U} L(u) = \bigcup_{\kappa \in \mathbb{F} \cup \infty} U(\kappa).$$

20.4.1 Theorem. *The subspaces $U(\kappa)$, where $\kappa \in \mathbb{F} \cup \infty$, form a regulus. All reguli arise in this way.*

Proof. Exercise. □

??Lemma 1.1 extends immediately to higher dimensions. Thus the pairwise skew subspaces $V(\sigma_i)$ lie on a regulus if and only if

$$(\sigma_2 - \sigma_4)^{-1}(\sigma_4 - \sigma_1) = \kappa(\sigma_2 - \sigma_3)^{-1}(\sigma_3 - \sigma_1).$$

(The proof of ??Lemma 1.1 actually works in all cases, not just in dimension four.) We will still call a spread regular if it contains all subspaces on the regulus generated by any three of its components. The translation plane determined by a regular spread is Pappian when the underlying field is finite.

We outline a proof of this, under the asumption that $|\mathbb{F}| > 2$. Our proof of ??Theorem 1.2 actually shows that the following is true. Suppose Σ determines a spread $\mathcal{S}$ of $V = U \oplus U$ containing $V(0)$ and $V(\infty)$ and, for any two components $V(\rho_1)$ and $V(\rho_2)$, all the subspaces on the regulus through $V(\infty)$, $V(\rho_1)$ and $V(\rho_2)$ belong to $\mathcal{S}$. Then $\Sigma \cup 0$ is a vector space over our underlying field $\mathbb{F}$. By ??Lemma 3.2.3, this implies that the translation plane π determined by the spread is $((\infty), V(\infty))$-transitive. Hence, if a spread is regular, p is a point on the line at infinity and ℓ is the line joining it to o then π is (p, ℓ)-transitive. If π is both (p, ℓ)- and (p, ℓ_∞)-transitive then it is (p, m)-transitive for all lines m through p. Hence p corresponds, in the dual plane, to a translation line. Accordingly each point on ℓ_∞ determines a translation line in the dual plane. By ??Theorem 3.4.2 the plane π is Moufang and, by ??Theorem 3.4.3, it must be Pappian.

Part III

Lines

Chapter 21

Lines and Bounds

We consider special sets of lines in real and complex space. The simplest to describe are *equiangular lines* in $\mathbb{R}^d$: this is a set of lines such that the angle between any two distinct lines is the same. If x and y are unit vectors spanning lines in $\mathbb{R}^d$, the angle between them is determined by $|\langle x, y\rangle|^2$. Note that this quantity does not change if we replace y (say) by $-y$. Now we can define a set of lines in $\mathbb{C}^d$ to be equiangular if there is a real constant α^2 such that, for any two unit vectors x and y spanning distinct lines in the set we have $|\langle x, y\rangle|^2 = \alpha^2$.

The other case of interest are sets of orthonormal bases of $\mathbb{R}^d$ or $\mathbb{C}^d$, such that for any two vectors x and y in distinct bases, the value of $|\langle x, y\rangle|^2$ is the same. A pair of orthonormal bases with this property is said to be *unbiased*, and in general a set of orthonormal bases is *mutually unbiased* if each pair is unbiased.

We will refer to $|\langle u, v\rangle|$ as the *angle* between the lines spanned by u and v. (This is an abuse of notation: in the real case, it actually the cosine of the angle between the lines and, in the complex case, it is not clear what an angle is.)

In the cases of interest to us, we will see that we can readily derive a good upper bound on the maximum size of a set of equiangular lines, or a set of mutually unbiased bases. The chief difficulty is to construct sets of lines realizing these bounds.

21.1 Projections

A line ℓ is a dimensional subspace of an inner product space V and hence can be represented by a basis, that is, any non-zero vector in ℓ. We can reduce redundancy by choosing a unit vector as a basis but even over the reals this still leaves a choice—between $-x$ and x—while over $\mathbb{C}$ a unit vector is only determined up to multiplication by a complex number of norm 1. (In quantum physics these are known as *phase factors*.)

We can eliminate redundancy by using projections. If U is a subspace of V with an orthonormal basis $u_1, \ldots, u_d$ then the matrix

$$P_U := \sum_{i=1}^{d} u_i u_i^*$$

represents orthogonal projection onto the subspace spanned by u. Hence

$$P = P^* = P^2$$

and U is the image of P_U and $U^\perp$ is its kernel. In particular

$$\operatorname{tr}(P_U) = \operatorname{rk}(P_U) = \dim(U).$$

The space of linear operators on V is an inner product space, with inner product

$$\langle A, B\rangle := \operatorname{tr}(A^*B) = \operatorname{sum}(\overline{A} \circ B).$$

If P and Q are projections then

$$\begin{aligned} \|P - Q\|^2 &= \operatorname{tr}((P-Q)^2) \\ &= \operatorname{tr}(P^2 + Q^2 - PQ - QP) \\ &= \operatorname{tr}(P) + \operatorname{tr}(Q) - 2\operatorname{tr}(PQ) \\ &= \operatorname{tr}(P) + \operatorname{tr}(Q) - 2\langle P, Q\rangle. \end{aligned}$$

If P and Q are projections onto subspaces of dimension d, it follows that

$$\|P - Q\|^2 = 2d - \langle P, Q\rangle;$$

if P and Q are projections onto lines spanned by unit vectors u and v respectively, then

$$P = uu^*, \quad Q = vv^*$$

and

$$\mathrm{tr}(PQ) = \mathrm{tr}(uu^*vv^*) = \mathrm{tr}(v^*uu^*v) = |u^*v|^2,$$

whence

$$\|P - Q\|^2 = 2 - 2|u^*v|^2.$$

Projections are Hermitian matrices, and the Hermitian matrices of order $d \times d$ form a real vector space of dimension d^2. Over $\mathbb{R}$ our projections are real symmetric matrices, and these form a real vector space of dimension $\binom{d+1}{2}$.

21.2 Equiangular Lines: The Absolute Bound

We derive upper bounds on the size of set of equiangular lines in $\mathbb{R}^d$ and $\mathbb{C}^d$. The observation is that a set of vectors in an inner product space is linearly independent if and only if the Gram matrix of the set is invertible.

Our first results provides the *absolute bound* on the maximum size of a set of equiangular lines.

21.2.1 Theorem. *The cardinality of a set of equiangular lines in $\mathbb{C}^d$ is at most d^2; in $\mathbb{R}^d$ it is at most $\binom{d}{2}$.*

Proof. Let $P_1, \ldots, P_m$ be the projections onto a set of equiangular lines with angle α and let G be the Gram matrix of these projections relative to the trace inner product. Then $Gi, i = 1$ for all i and, if $i \neq j$. then $G_{i,j} = \alpha^2$. Therefore

$$G = (1 - \alpha^2)I + \alpha^2 J$$

and, since $\alpha^2 < 1$, it follows that G is invertible. This implies that the matrices $P_1, \ldots, P_m$ are linearly independent.

Over $\mathbb{C}$, projections are Hermitian and the space of complex Hermitian matrices has dimension d^2 (over $\mathbb{R}$). Over $\mathbb{R}$, projections are symmetric and the space of real symmetric matrices has dimension $\binom{d}{2}$. □

21.2.2 Lemma. *Assume $P_1, \ldots, P_m$ are projections onto a set of equiangular lines in $\mathbb{C}^d$ or $\mathbb{R}^d$. If the absolute bound is tight, then*

$$\sum_r P_r = \frac{m}{d} I$$

Proof. If the absolute bound is tight, the projections $P_1, \ldots, P_m$ are a basis for the space of Hermitian/symmetric matrices. So there are constants c_r such that

$$I = \sum_r c_r P_r.$$

Assume that angle between distinct lines is α. Then

$$1 = \langle I, P_s \rangle = \sum_r c_r \langle P_s, P_r \rangle = c_s + \sum_{r \neq s} c_r \alpha^2 = (1 - \alpha^2) c_s + \sum_r c_r.$$

From this we see first that the coefficients c_r are all equal, and hence they are all equal to d/n. □

A set of unit vectors $x_1, \ldots, x_m$ such that

$$\sum_r x_r x_r^* = \frac{m}{d} I$$

is known as a *tight frame*. A set of unit vectors that is equiangular and a tight frame is an *equiangular tight frame*. (The size of an equiangular tight frame need not meet the absolute bound, as we will see.)

The six diagonals of the icosahedron in $\mathbb{R}^3$ form an equiangular tight frame, so at least the real absolute bound is tight once. (We will say more later.)

21.2.3 Lemma. *If there is an equiangular tight frame in $\mathbb{C}^d$ with angle α, then $\alpha^2 = (d+1)^{-1}$; In $\mathbb{R}^d$ we have $\alpha^2 = (d+2)^{-1}$.*

Proof. If

$$\sum_r P_r = \frac{m}{d} I$$

then

$$P_1 + \sum_{r>1} P_1 P_r = \frac{m}{d} P_1$$

and taking traces yields

$$1 + (m-1)\alpha^2 = \frac{m}{d}.$$

Therefore

$$\alpha^2 = \frac{m-d}{d(m-1)}$$

and, on substituting in the correct value for m, we get the stated values for α. □

21.3 Equiangular Lines: The Relative Bound

We derive what we call the *relative bound* on the size of an equiangular set of lines; this bound depends on the angle as well as the dimension.

21.3.1 Theorem. *If there is an equiangular set of n lines in $\mathbb{C}^d$ or $\mathbb{R}^d$ with angle α and $d\alpha^2 < 1$, then*

$$n \leq \frac{d - d\alpha^2}{1 - d\alpha^2}.$$

Equality holds if and only if the lines form an equiangular tight frame.

Proof. Let $P_1, \ldots, P_m$ be projections onto equiangular lines with angle α and set

$$S = \frac{m}{d} I - \sum_r P_r.$$

Then

$$\begin{aligned} 0 \leq \langle S, S \rangle &= \frac{m^2}{d} - 2\frac{m^2}{d} + m + m(m-1)\alpha^2 \\ &= m\left(-\frac{m}{d} + 1 + (m-1)\alpha^2\right) \end{aligned}$$

and consequently

$$\left(\frac{1}{d} - \alpha^2\right) m \leq 1 - \alpha^2;$$

this yields our bound. Equality holds if and only if $S = 0$. □

Note that if $\alpha^2 = (d+1)^{-1}$, we recover the absolute bound for complex equiangular lines and if $\alpha^2 = (d+2)^{-1}$ we get the real absolute bound.

21.4 Type-II Matrices

We denote the Schur inverse of a matrix W, if it exists, by $W^{(-)}$. An $n \times n$ complex matrix W is a *type-II* matrix if it is Schur invertible and

$$WW^{(-)T} = nI.$$

Any Hadamard matrix is an example. We say that a complex matrix is *flat* if its entries all have the same absolute value.

21.4.1 Lemma. *Any two of the following conditions on a square complex matrix W imply the third:*

(a) W is type-II.

(b) W is flat.

(c) W is a non-zero scalar multiple of a unitary matrix. □

We leave the proof of this an exercise. You should also show that the Kronecker product of two type-II matrices is type-II. The next result provides a class of examples of type-II matrices that are not flat in general.

21.4.2 Lemma (Chan and Godsil)**.** *Let N be a square 01-matrix and suppose $W = aJ + (b-a)N$, where $a \neq \pm b$. Then W is type-II if and only if N is the incidence matrix of a symmetric design.* □

We introduce another class of type-II matrices. We say that an $n \times n$ matrix C is a *generalized conference matrix* if:

(a) C is Hermitian.

(b) $C \circ I = 0$.

(c) If $i \neq j$, then $|C_{i,j}| = 1$.

(d) The minimal polynomial of C is quadratic

It is easy to check that symmetric conference matrices are generalized conference matrices. If C is a skew-symmetric conference matrix, then iC is Hermitian and is hence a generalized conference matrix. In both these cases the minimal polynomial is $t^2 - n + 1$ and $C + iI$ respectively $i(C + I)$ is a flat type-II matrix.

If the minimal polynomial of C is $t^2 + \beta t + \gamma$, then

$$C^2 + \beta C + \gamma I = 0$$

and so $C^2 \circ I - \gamma I$. It follows that $\gamma = n - 1$.

21.4.3 Theorem (Chan and Godsil)**.** *If C is a generalized conference matrix with minimal polynomial $t^2 - \beta t + n - 1$ and $z + z^{-1} + \beta = 0$, then $zI + C$ is type-II.* □

We leave the proof of this as an exercise—it is a straightforward verification. In [**??**], it is shown that if W is a Hermitian type-II matrix with constant diagonal, then it arises from a generalized conference matrix.

In the next section we establish a connection between generalized conference matrices and equiangular tight frames.

21.5 Type-II Matrices from Equiangular Tight Frames

We have used the Gram matrix of a set of projections onto equiangular lines, now we turn to a different Gram matrix. Assume $x_1, \ldots, x_m$ are a set of unit vectors in dimension d, spanning a set of equiangular lines with angle α. Then we can write their Gram matrix G as

$$G = I + \alpha^2 S$$

where S is Hermitian with zero diagonal and all off-diagonal entries have absolute value 1. We call it the *Seidel matrix* of the set of lines. Let U be the $n \times d$ matrix with i-th row u_i^*. Then

$$G = UU^*$$

and

$$U^*U = \sum_r x_r x_r^*.$$

21.5.1 Theorem. *A set of equiangular lines forms a tight frame if and only if its Seidel matrix is a generalized conference matrix.*

Proof.

21.6 Lines with Few Angles from Group Matrices

The *degree set* of a set of lines with projections $P_1, \ldots, P_m$ is the set

$$\{\langle P_r, P_s\rangle : r \neq s\}.$$

Let Γ be an abelian group. We use Γ^* to denote its group of characters.

21.6.1 Theorem (Godsil and Roy)**.** *Let X be a connected k-regular bipartite graph with adjacency matrix*

$$A = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix},$$

where B is a group matrix for an abelian group Γ of order n. Let u be a vertex in X and if ψ is a character of Γ, let ψ_u denote the restriction of ψ to the neighbourhood of u. Then the degree set of the lines in $\mathbb{C}^k$ spanned by the vectors ψ_u consists of the numbers θ/k, where θ runs over the non-negative eigenvalues of X not equal to k.

Chapter 22

Real Lines

The first definition is easy enough: a set of lines in $\mathbb{R}^n$ is *equiangular* if the angle between any two distinct lines is the same. The simplest example would be the coordinate axes in $\mathbb{R}^d$, which gives us a set of size d. The first problem is to determine the maximum size of a set of equiangular lines in $\mathbb{R}^d$.

22.1 Projections

A line ℓ is a 1-dimensional subspace of V and hence can be represented by a basis, that is, any non-zero vector in ℓ. We can reduce redundancy by choosing a unit vector as a basis but even over the reals this still leaves a choice—between $-x$ and x—while over $\mathbb{C}$ a unit vector is only determined up to multiplication by a complex number of norm 1. (In quantum physics these are known as *phase factors*.)

We can eliminate redundancy by using projections. If U is a subspace of V with an orthonormal basis $u_1, \ldots, u_d$ then the matrix

$$P_U := \sum_{i=1}^{d} u_i u_i^*$$

represents orthogonal projection onto. Hence

$$P = P^* = P^2$$

and U is the image of P_U and $U^\perp$ is its kernel. In particular

$$\operatorname{tr}(P_U) = \operatorname{rk}(P_U) = \dim(U).$$

The space of linear operators on V is an inner product space, with inner product

$$\langle A, B\rangle := \operatorname{tr}(A^*B) = \operatorname{sum}(\overline{A} \circ B).$$

If P and Q are projections then

$$\begin{aligned}\|P-Q\|^2 &= \operatorname{tr}((P-Q)^2)\\ &= \operatorname{tr}(P^2+Q^2-PQ-QP)\\ &= \operatorname{tr}(P)+\operatorname{tr}(Q)-2\operatorname{tr}(PQ)\\ &= \operatorname{tr}(P)+\operatorname{tr}(Q)-2\langle P,Q\rangle.\end{aligned}$$

If P and Q are projections onto subspaces of dimension d, it follows that

$$\|P-Q\|^2 = 2d - \langle P, Q\rangle;$$

if P and Q are projections onto lines spanned by unit vectors u and v respectively then

$$P = uu^*, \quad Q = vv^*$$

and

$$\operatorname{tr}(PQ) = \operatorname{tr}(uu^*vv^*) = \operatorname{tr}(v^*uu^*v) = |u^*v|^2$$

whence

$$\|P-Q\|^2 = 2 - 2|u^*v|^2.$$

If we are working over $\mathbb{R}$ then $|u^*v|$ is the cosine of the angle between the lines spanned by u and v. Hence we will call $|u^*v|^2$ a *squared cosine*, even over $\mathbb{C}$.

Projections are Hermitian matrices, and the Hermitian matrices of order $d \times d$ form a real vector space of dimension d^2. Over $\mathbb{R}$ our projections are real symmetric matrices, and these form a real vector space of dimension $\binom{d+1}{2}$.

22.2 Equiangular Lines

We begin by a deriving sharp upper bound on the size of a set of equiangular lines in $\mathbb{R}^d$. The first step is another representation of lines. Suppose x is a non-zero vector. Then the matrix

$$X = \frac{1}{\langle x, x\rangle} xx^T$$

is symmetric and idempotent and its image is the line spanned by x. Thus X represents orthogonal projection onto the line spanned by x. Note that if we replace x by c, where $c \neq 0$, the matrix X does not change. In particular x and $-x$ give rise to the same matrix X. Thus X represents our line and does not depend on the choice the basis of the line.

Further, suppose x_i and x_j are unit vectors and

$$X_i := x_i x_i^T, \quad X_j = x_j x_j^T.$$

Then

$$X_i X_j = \langle x_i, x_j \rangle x_i x_j^T$$

and

$$\langle X_i, X_j \rangle = \operatorname{tr}(X_i X_j) = \langle x_i, x_j \rangle^2.$$

Thus $\langle X_i, X_j \rangle$ is the squared cosine of the angle between the lines spanned by x_i and x_j. Also $\langle X_i, X_i \rangle = 1$.

22.2.1 Theorem. *A set of equiangular lines in $\mathbb{R}^d$ has size at most $\binom{d+1}{2}$.*

Proof. Suppose our lines are spanned by vectors $x_1, \ldots, x_n$, with corresponding projections $X_1, \ldots, X_n$, and that the square cosine is γ. We prove that the matrices $X_1, \ldots, X_n$ are linearly independent. Since these matrices lie in the real vector space of $d \times d$ symmetric matrices, which has dimension

$$\binom{d+1}{2},$$

the theorem follows immediately.

Assume that $c_1, \ldots, c_n$ are scalars and

$$0 = \sum_{i=1}^{n} c_i X_i.$$

Take the inner product of each side with X_r. Then

$$\begin{aligned} 0 &= c_r + \sum_{i \neq r} c_i \gamma \\ &= (1 - \gamma) c_r + \gamma \sum_i c_i. \end{aligned}$$

Since this holds for $r = 1, \ldots, n$, we see that c_r is independent of r. Therefore we must have

$$0 = \sum_{i=1}^{n} X_i,$$

but the trace of the right side is n, and so it cannot be zero. We conclude that the matrices X_i are linearly independent. □

The above bound on the size of an equiangular set of lines is known as the *absolute bound*. You may convince yourself that it is tight in $\mathbb{R}^2$. (We will consider the question of tightness in more detail later.) Note that if we have an equiangular set of n lines in $\mathbb{R}^d$, the intersection of these lines gives us a set of $2n$ points, namely the $2n$ unit vectors that span the lines. If x_i and x_j are two of these vectors, then

$$\|x_i - x_j\|^2 = 2 - 2\langle x_i, x_j \rangle \geq 2 - 2\sqrt{\gamma}$$

Using this and some spherical geometry, we could derive an upper bound on the size of our set. However the resulting bound depends on γ and is exponential in d.

22.3 The Relative Bound

We derive a second bound on the size of an equiangular set of lines. This bound depends both on d and the squared cosine γ. (It will also be easier to give examples where it is tight.)

Suppose $X_1, \ldots, X_n$ are the projections onto a set of equiangular lines in $\mathbb{R}^d$ with squared cosine γ. If the number of lines meets the absolute bound, then these projections span the vector space of $d \times d$ symmetric matrices, whence there are scalars c_i such that

$$I = \sum_i c_i X_i.$$

If we take the inner product of each side with X_r, we find that

$$1 = (1 - \gamma)c_r + \gamma \sum_i c_i$$

As before this implies that c_r is independent of r. Comparing traces, we conclude that, if the absolute bound holds, then

$$\sum_i X_i = \frac{n}{d} I.$$

This may motivate the following. We forget the absolute bound and compute

$$\left\langle \nu I - \sum_i X_i, \nu I - \sum_i X_i \right\rangle = \nu^2 d - 2\nu n + n + (n^2 - n)\gamma$$

Here the left side is nonnegative for any choice of ν, so we substitute n/d for ν in the right side and deduce that

$$\begin{aligned} 0 &\leq \frac{n^2}{d} - 2\frac{n^2}{d} + n + (n^2 - n)\gamma \\ &= \frac{n}{d}(-n + d + d(n-1)\gamma) \end{aligned}$$

and consequently

$$n(1 - d\gamma) \leq d - d\gamma.$$

If $d\gamma < 1$, we conclude that the following *relative bound* holds.

22.3.1 Theorem. *If there is an equiangular set of n lines in $\mathbb{R}^d$ with squared cosine γ, then*

$$n \leq \frac{d - d\gamma}{1 - d\gamma}$$

and, if equality holds and $X_1, \ldots, X_n$ are the projections onto the lines, then

$$\sum_i X_i = \frac{n}{d} I. \qquad \square$$

If equality holds, we will see that the possible values for γ are quite restricted. If

$$\sum_i X_i = \frac{n}{d}$$

then taking the inner product of each side with X_1 yields

$$1 - \gamma + n\gamma = \frac{n}{d}$$

and consequently

$$\gamma = \frac{n - d}{d(n-1)}.$$

In particular, if we have an equiangular set of $\binom{d+1}{2}$ lines in $\mathbb{R}^d$ with projections $X_1, \ldots, X_n$, then

$$\gamma = \frac{1}{d+2}.$$

A set of unit vectors $x_1, \ldots, x_n$ forms a *tight frame* in $\mathbb{R}^d$ if

$$\sum_i x_i x_i^T = \frac{n}{d} I.$$

22.4 Gram Matrices

Suppose $x_1, \ldots, x_n$ is a set of unit vectors in $\mathbb{R}^d$ that span a set of equiangular lines with squared cosine γ, and let G be their Gram matrix. Then we may write

$$G = I + \sqrt{\gamma} S,$$

where S is a symmetric matrix with all diagonal entries zero, and all off-diagonal entries equal to ± 1. We call S the *Seidel matrix* of the set of vectors. Further

$$\frac{1}{2}(J - I + S)$$

is the adjacency matrix of a graph. Since the lines are determined up to an orthogonal transformation by the gram matrix, it follows that each equiangular set of lines is determined by a graph. (The correspondence is many-to-one, since we may replace x_i by $-x_i$ without changing the set of lines. The leads to the concept of *switching classes* of graphs, but we do not go into this now.)

The next result will be the key to our analysis.

22.4.1 Lemma. *Suppose $x_1, \ldots, x_n$ is a set of unit vectors in $\mathbb{R}^d$ that span a set of equiangular lines with squared cosine γ and let G be their Gram matrix. If the relative bound holds with equality, then*

$$G^2 = \frac{n}{d} G.$$

Proof. Let $X_i = x_i x_i^T$. If the relative bound is tight, then

$$\sum_i X_i = \frac{n}{d} I.$$

Let U be the $d \times n$ matrix with $x_1, \ldots, x_n$ as its columns. Then

$$\sum_i X_i = U U^T$$

and $G = U^T U$. Hence

$$G^2 = U^T(UU^T)U = \frac{n}{d}U^T U = \frac{n}{d}G. \qquad \square$$

It follows that the minimal polynomial of G is

$$t^2 - \frac{n}{d}t$$

and therefore the eigenvalues of G are n/d (with multiplicity d) and 0 (with multiplicity $n-d$). If

$$S = \frac{1}{\sqrt{\gamma}}(G - I)$$

then the eigenvalues of S are

$$\frac{n-d}{d\sqrt{\gamma}}, \; -\frac{1}{\sqrt{\gamma}}$$

with respective multiplicities d and $n-d$. A symmetric matrix with one eigenvalue is a scalar multiple of I; we have found that if there is an equiangular set of lines meeting the relative bound, then the Seidel matrix S is a symmetric matrix with only two eigenvalues. Note that the procedure is reversible: given a Seidel matrix with only two eigenvalues, we can construct a set of equiangular lines with size meeting the relative bound.

If S is a Seidel matrix with exactly two eigenvalues α and β, then

$$0 = (S - \alpha I)(S - \beta I) = S^2 - (\alpha + \beta)S + \alpha\beta I.$$

Since the diagonal of S is zero, it follows that each diagonal entries of S^2 is equal to $-\alpha\beta$. On the other hand since the off-diagonal entries of S are all $\pm$, each each diagonal entry of S^2 is equal to $n-1$. Thus we see that the product of the eigenvalues of S is $1-n$. Hence the eigenvalues of S are $(n-1)\sqrt{\gamma}$ and $-1/\sqrt{\gamma}$.

22.5 Number Theory

Suppose $\mathbb{F}$ and $\mathbb{E}$ are fields and $\mathbb{F} \leq \mathbb{E}$. If $a \in \mathbb{E}$, the minimum polynomial of a over $\mathbb{F}$ is the monic polynomial ψ of least degree with coefficients in $\mathbb{F}$ such that $\psi(a) = 0$, if it exists. We will only be concerned with cases where

$\mathbb{F} = \mathbb{Q}$ and $\mathbb{E} = \mathbb{C}$. An element which does not have a minimal polynomial is transendental, otherwise it is algebraic. The elements of $\mathbb{C}$ whose minimal polynomial over $\mathbb{Q}$ have integer coefficients are called *algebraic integers*; these form a ring. The minimal polynomial of a over $\mathbb{F}$ is irreducible over $\mathbb{F}$. Two elements of $\mathbb{E}$ are algebraic conjugates over $\mathbb{F}$ if they have the same minimal polynomial.

Note that $\mathbb{E}$ is a vector space over $\mathbb{F}$, we denote its dimension by $|\mathbb{E} : \mathbb{F}|$ and, if it is finite, we may say that $\mathbb{E}$ is an extension of $\mathbb{F}$ with degree equal to $\mathbb{E} : \mathbb{F}|$. A quadratic extension is an extension of degree two. If $a \in \mathbb{E}$, then the set $\mathbb{F}[a]$ of all polynomials in a with coefficients from $\mathbb{F}$ is a vector space over $\mathbb{F}$; its dimension is the degree of the minimal polynomial of a.

Suppose $\phi(t)$ is the characteristic polynomial of the integer matrix A. If λ is an eigenvalue of A, then$\phi(\lambda) = 0$ and it follows that the minimal polynomial of a divides ϕ. Hence each zero of the minimal polynomial, that is, each algebraic conjugate of λ, is an eigenvalue of A. Further all algebraic conjugates of λ will have the same algebraic multiplicity.

22.5.1 Lemma. *If there is an equiangular set of n lines in $\mathbb{R}^d$ with squared cosine γ such that the relative bound holds, then either $1/\sqrt{\gamma}$ is an integer, or $n = 2d$ and $1/\sqrt{\gamma}$ lies in a quadratic extension of the rationals.*

Proof. Since S is an integer matrix, its eigenvalues are algebraic integers. Further if λ is an eigenvalue of S, then all its algebraic conjugates are eigenvalues of S with multiplicities equal to the multiplicity of λ. Since S has exactly two eigenvalues with multiplicities $n-d$ and d we see that either λ is an integer, or $n = 2d$ and λ is a quadratic irrational. Since $-1/\sqrt{\gamma}$ is an eigenvalue of S, the second claim follows. □

We have seen that if we have an equiangular set of $\binom{d+1}{2}$ lines in $\mathbb{R}^d$, then $\gamma = (d+2)^{-1}$. Hence we have the following.

22.5.2 Corollary. *If there is an equiangular set of lines in R^d meeting the absolute bound and $d \geq 4$, then $d+2$ is the square of an integer.* □

We will see later that $d+2$ must actually be the square of an odd integer. Examples of sets of lines meeting the absolute bound are known when $d = 2$, 3, 7 or 23 (and we will present them later). No other examples are known.

22.6 Switching

If X is a graph then

$$S = A(X) - A(\overline{X})$$

is a Seidel matrix and so, if X has v vertices and least eigenvalue τ with multiplicity m, then

$$S - \gamma I$$

is the Gram matrix of an equiangular set of v lines in $\mathbb{R}^{v-m}$ with squared cosine γ^{-2}. Suppose D is a $v \times v$ diagonal matrix with diagonal entries ± 1. Then $D = D^{-1}$ and so DSD is a Seidel matrix which is similar to S. As far as lines are concerned, replacing S by DSD is equivalent to multiplying some of the spanning unit vectors -1, and so geometrically nothing interesting is happening. However DSD is the Seidel matrix of some graph Y, and we want to determine the relation between X and Y.

If $\sigma \subseteq V(X)$, we define X^σ to be the graph we get from X by complementing the edges that join vertices in σ to vertices not in σ. If $\bar{\sigma}$ denotes the complement of σ, then in set theoretic terms $E(X^\sigma)$ is the symmetric difference of $E(X)$ and the edge set of the complete bipartite graph with bipartition

$$(\sigma, \bar{\sigma})$$

We say that X^σ is obtained by *switching* about the subset σ. Note that switching twice about σ restores X to itself, and that

$$X^\sigma = X^{\bar{\sigma}}.$$

If D is the $v \times v$ diagonal matrix such that $D_{i,i} = -1$ if $i \in \sigma$ and $D_{i,i} = 1$ if $i \notin \sigma$, then

$$DS(X)D = S(X^\sigma).$$

This reconciles the graph theory and the linear algebra.

It is not hard to show that any sequence of switchings on subsets of X can be realised by switching on a single subset. So we say graphs X and Y are *switching equivalent* if Y is isomorphic to X^σ for some σ, and the graphs that are switching equivalent to X form its *switching class*. If σ is the neighborhood of a vertex v in X, then v is an isolated vertex in X^σ; in this case we say that X^σ is obtained by *switching off* v. In 22.10 we introduce the graph of a set of equiangular lines; this determines the switching class of X. (More precisely, it reduces switching equivalence of graphs on v vertices to isomorphism of certain graphs on $2v$ vertices.)

22.7 Paley Graphs

Let $\mathbb{F}$ be a finite field of order q, where $q \equiv 1$ modulo four. The Paley graph on q vertices has $\mathbb{F}$ as its vertex set, and two field elements are adjacent if and only if their difference is a non-zero square in $\mathbb{F}$. (The condition on q assures that we obtain a graph rather than a directed graph. The 5-cycle is the Paley graph associated to the field of order 5.

A Paley graph is self-complementary and is strongly regular with parameters

$$\left(q, \frac{q-1}{2}; \frac{q-5}{4}, \frac{q-1}{4}\right).$$

Its eigenvalues are its valency (with multiplicity 1) and

$$\frac{1}{2}(1 \pm \sqrt{q}).$$

each with multiplicity $(q-1)/2$.

22.7.1 Lemma. *If X is a Paley graph on q vertices and $S = S(X \cup K_1)$, then $S^2 = qI$.*

Proof. Exercise. □

Since $\mathrm{tr}(S) = 0$, the eigenvalues $\pm\sqrt{q}$ each have multiplicity $(q+1)/2$. Hence

$$S + \sqrt{q}I$$

is the Gram matrix of a equiangular set of $q+1$ lines in $\mathbb{R}^{(q+1)/2}$.

In particular, the Paley graph on five vertices provides us with a set of six equiangular lines in $\mathbb{R}^3$. This realizes the absolute bound (and provides a construction of the icosahedron).

A $v \times v$ matrix C with zero diagonal and entries ± 1 off the diagonal is a *conference matrix* if

$$C^T C = (v-1)I.$$

The Seidel matrices we have just constructed are symmetric conference matrices.

22.8 A Spherical 2-Design

Suppose we have a set of n equiangular lines in $\mathbb{R}^d$ with squared cosine γ. We may assume without loss that one of the lines is spanned by the first standard basis vector e_1, and then we can choose unit vectors $x_2, \ldots, x_n$ spanning the remaining $n-1$ lines so that

$$\langle e_1, x_i \rangle = \sqrt{\gamma}.$$

This means that each vector x_i can be written as

$$x_i = \begin{pmatrix} \sqrt{\gamma} \\ y_i \end{pmatrix}$$

where $\|y_i\| = \sqrt{1-\gamma}$. Hence the projection onto the line spanned by x_i has the form

$$\begin{pmatrix} \gamma & \sqrt{\gamma} y_i^T \\ \sqrt{\gamma} y_i & y_i y_i^T \end{pmatrix}.$$

Now assume that the relative bound is tight. If $X_1, \ldots, X_n$ denote the projections onto our lines, then

$$\sum_{i=1}^{n} X_i = \frac{n}{d} I.$$

If we let z_i denote the unit vector $(1-\gamma)^{-1/2} y_i$, then we have

$$\sum_i z_i = 0, \qquad \sum_i z_i z_i^T = \frac{n}{d - d\gamma} I.$$

It follows that the vectors z_i provide an example of what we will come to call a *spherical 2-design*. Now we simply show that these vectors determine a strongly regular graph on $n-1$ vertices.

The first step is to note that since

$$\gamma + y_i^T y_j = x_i^T x_j = \pm\sqrt{\gamma}$$

we have

$$z_i^T z_j = \frac{\pm\sqrt{\gamma} - \gamma}{1 - \gamma}.$$

We define a graph G with the vectors z_i as its vertices, where two distinct vectors are adjacent if their inner product is positive. Let Z denote the

$(d-1)\times(n-1)$ matrix with the vectors z_i as its columns. If $A := A(G)$, then

$$\begin{aligned} Z^TZ &= I + \frac{\sqrt{\gamma}-\gamma}{1-\gamma}A - \frac{\sqrt{\gamma}+\gamma}{1-\gamma}(J-I-A) \\ &= \frac{1+\sqrt{\gamma}}{1-\gamma}I + \frac{2\sqrt{\gamma}}{1-\gamma}A - \frac{\sqrt{\gamma}+\gamma}{1-\gamma}J. \end{aligned} \tag{22.8.1}$$

On the other hand

$$ZZ^T = \sum_i z_i z_i^T = \frac{n}{d-d\gamma}I$$

and therefore

$$(Z^TZ)^2 = \frac{n}{d-d\gamma}Z^TZ.$$

This implies that Z^TZ has exactly two eigenvalues, and from (22.8.1) it follows that A has exactly three eigenvalues. Since $\sum z_i = 0$, we see that

$$JZ^TZ = Z^TZJ = 0,$$

and therefore G is regular. Thus G is a regular graph with three eigenvalues, and therefore it is strongly regular. (A strongly regular graph can arise in this way if and only if $k = 2c$.)

You are invited to show that if z is an eigenvector of A that is orthogonal to $\mathbf{1}$, then its eigenvalue is one of

$$\frac{1}{2}\left(\frac{n-d}{d\sqrt{\gamma}}-1\right) = \frac{1}{2}[(n-1)\sqrt{\gamma}-1], \qquad \frac{1}{2}\left(-\frac{1}{\sqrt{\gamma}}-1\right).$$

If we compare these with the eigenvalues of the Seidel matrix, we deduce that if the eigenvalues of the Seidel matrix are integers, they must be odd integers.

For any strongly regular graph, $c-k = \theta\tau$ and as $k = 2c$ it follows that $k = -2\theta\tau$. Hence the valency of G is

$$k = \frac{1}{2\sqrt{\gamma}}((n-1)\sqrt{\gamma}-1))(1+\sqrt{\gamma}).$$

22.9 An Example

We construct a set of 28 vectors $x_{i,j}$ in $\mathbb{R}^8$ by defining $x_{i,j}$ to be the vector with i-th and j-th entries equal to 3, and all other entries equal to -1. The entries of each of these vectors sum to zero, and so they span a set of lines in $\mathbb{R}^7$ with squared cosine $1/9$. Since

$$28 = \binom{8}{2},$$

we have equality in the absolute bound. Choose x_1 to be the vector with first two entries equal to 3 (which is not a unit vector but that will not matter). The neighbors of x_1 in the graph of the lines consists of the 27 vectors of the form $\pm x_{i,j}$ with positive inner product with x_1. This set of vectors consists of the 12 vectors with first or second entry equal to 3, and the fifteen vectors with first two entries equal to 1.

The eigenvalues of the Seidel matrix are 9 and -3, and the eigenvalues of the neighborhood in the two-graph are 4 and -2. The valency is 16. If the eigenvalues of the neighborhood are k, θ and τ, then

$$(t-\theta)(t-\tau) = t^2 - (a-c)t - (k-c).$$

Hence we have

$$c = k + \theta\tau, \quad a = c + \theta + \tau = k + \theta\tau + \theta + \tau$$

and for the graph at hand

$$c = 8, \quad a = 10.$$

22.10 Graphs from Equiangular Lines

Let $x_1, \ldots, x_n$ be a set of n unit vectors in R^d, spanning a set of equiangular lines with squared cosine γ. The *graph* of this set of lines has the $2n$ vectors $\pm x_i$ as its vertices, and two such vectors are deemed to be adjacent if their inner product is $\sqrt{\gamma}$. Thus its vertex set is partitioned into n pairs

$$\{x_i, -x_i\}$$

and if $j \neq i$ then x_j is adjacent to exactly one of the vectors x_i and $-x_i$. So the subgraph induced by two pairs is isomorphic to $2K_2$.

22.10.1 Lemma. *If Y is the graph of a set of equiangular lines, then we may write $A(Y)$ in the form*

$$A(Y) = \begin{pmatrix} A(X) & A(\overline{X}) \\ A(\overline{X}) & A(X) \end{pmatrix}$$

where $A(X) - A(\overline{X})$ is the Seidel matrix of the set of lines. □

22.10.2 Corollary. *Suppose Y is the graph of a set of n equiangular lines. Then*

$$\phi(A(Y), t) = \phi(S, t)\phi(K_n, t).$$

Proof. If

$$A(Y) = \begin{pmatrix} A & \overline{A} \\ \overline{A} & A \end{pmatrix}$$

where $A = A(X)$ for some graph X, then

$$\begin{pmatrix} I & 0 \\ I & I \end{pmatrix} \begin{pmatrix} A & \overline{A} \\ \overline{A} & A \end{pmatrix} \begin{pmatrix} I & 0 \\ -I & I \end{pmatrix} = \begin{pmatrix} A - \overline{A} & \overline{A} \\ 0 & A + \overline{A} \end{pmatrix}.$$ □

As an exercise, you may prove that if the graph of an equiangular set of n lines is either connected with diameter three, or is isomorphic to $2K_n$.

If u, v and w are distinct vertices in the graph of an equiangular set of lines and

$$\mathrm{dist}(u, v) = \mathrm{dist}(u, w) = 3$$

then

$$N(v) \setminus w = N(w) \setminus c,$$

since $N(v)$ and $N(w)$ are sets of size $n - 1$ disjoint from $u \cup N(u)$.

22.10.3 Theorem. *If $\mathcal{L}$ is an equiangular set of lines in $\mathbb{R}^d$ that meets the relative bound, then its graph is an antipodal distance-regular graph of diameter three, and the neighbourhood of any vertex is strongly regular.*

Proof. We have already proved the second claim, in 22.8. Given this it is easy to show that the graph is distance regular and antipodal with diameter three. Do it. □

The graph of a set of equiangular lines is sometimes called a *two-graph*; we say a two-graph is *regular* if each neighborhood is regular. A two-graph is regular if and only if the size of corresponding set of lines meets the relative bound.

22.10.4 Theorem. *If Y is a two-graph, the following are equivalent:*

(a) Y is distance regular.

(b) The neighborhood of each vertex of Y is regular.

(c) The neighborhood of each vertex of Y is strongly regular.

(d) The neighborhood of some vertex is strongly regular with $k = 2c$. □

Chapter 23

Complex Lines

We investigate the complex analogs of the results in the previous chapter.

23.1 The Absolute Bound

If x is a non-zero vector in $\mathbb{C}^d$ then the matrix

$$\frac{1}{x^*x}xx^*$$

represents orthogonal projection onto the line spanned by x. This is a Hermitian matrix with rank one. If X and Y are projections onto complex lines, we define the inner product

$$\operatorname{tr}(X^*Y)$$

to be the squared cosine of the angle between the two lines. If x and y are unit vectors and $X = xx^*$ and $Y = yy^*$, then

$$\operatorname{tr}(X^*Y) = \langle x, y\rangle\langle y, x\rangle = |\langle x, y\rangle|^2.$$

23.1.1 Theorem. *A set of equiangular lines in $\mathbb{C}^d$ has size at most d^2.*

Proof. Suppose $X_1, \ldots, X_n$ are the projections onto an equiangular set of n lines in $\mathbb{C}^d$ with squared cosine γ. We show that these projections form a linearly independent set in the vector space of Hermitian $d \times d$ matrices, and deduce the bound from this.

Assume that we have real scalars $c_1, \ldots, c_n$ such that

$$0 = \sum_i c_i X_i.$$

If we take the inner product of both sides with X_r, on the left, we get

$$0 = (1-\gamma)_r + \sum_i c_i \gamma,$$

from which we deduce that c_r is independent of r and hence that

$$0 = \sum_i X_i.$$

Since the trace of the right side is n, we have a contradiction and so we conclude that $X_1, \ldots, X_n$ is linearly independent.

The set of $d \times d$ Hermitian matrices is a **real** vector space with dimension d^2, and therefore $n \leq d^2$ as asserted. □

23.2 The Relative Bound

Physicists are only interested in equiangular sets of size d^2; we will consider a broader class of problems.

Suppose $X_1, \ldots, X_n$ are the projections onto a set of n equiangular lines with squared cosine γ, and that there are scalars $c_1, \ldots, c_n$ such that

$$I = \sum_i c_i X_i.$$

Then taking the inner product with X_r as before, we deduce that c_r is independent of r, and hence that

$$\sum_i X_i = \frac{n}{d} I.$$

It follows that

$$1 - \gamma + n\gamma = \frac{n}{d}$$

and so

$$\gamma = \frac{n-d}{d(n-1)}$$

When $n = d^2$, this yields that

$$\gamma = \frac{1}{d+1}.$$

23.2.1 Theorem. *If there is an equiangular set of n lines in $\mathbb{C}^d$ with squared cosine γ and $d\gamma < 1$, then*

$$n \leq \frac{d - d\gamma}{1 - d\gamma}$$

and, if equality holds and $X_1, \ldots, X_n$ are the projections onto the lines, then

$$\sum_i X_i = \frac{n}{d} I.$$

Proof. Exercise. □

23.3 Gram Matrices

The Gram matrices of sets of equiangular lines in $\mathbb{C}^d$ do not lead to graphs in general, but they still have some interesting properties.

Suppose $x_1, \ldots, x_n$ are unit vectors spanning a set of equiangular lines in $\mathbb{C}^d$ with squared cosine γ, let G be their Gram matrix and let S be the matrix defined by

$$G = I + \sqrt{\gamma} S.$$

Thus S is a Hermitian matrix with zero diagonal and with all off-diagonal entries having absolute value 1.

Assume now that the relative bound is tight, and let Z be the $d \times n$ matrix with the vectors $x_1, \ldots, x_n$ as its columns. Then

$$G = Z^T Z$$

and

$$Z Z^T = \frac{n}{d} I,$$

whence

$$G^2 = \frac{n}{d} G.$$

Thus the eigenvalues of G are 0 (with multiplicity $n - d$) and n/d (with multiplicity d) and therefore the eigenvalues of S are

$$-\frac{1}{\sqrt{\gamma}}, \quad \frac{n - d}{d\sqrt{\gamma}}$$

with respective multiplicities $n - d$ and d.

In general the entries of S are not integers and thus we cannot argue, as we did in the real case, that $\gamma^{-1/2}$ must be an integer.

23.4 Type-II Matrices

We use $A \circ B$ to denote the Schur product of two matrices with the same order. If $A \circ B = J$, we say that B is the *Schur inverse* of A, and write

$$B = A^{(-)}.$$

A $v \times v$ complex matrix W is a *type-II matrix* if

$$WW^{(-)T} = vI.$$

Note that if W is any Schur invertible $v \times v$ matrix, then the diagonal entries of $WW^{(-)T}$ are all equal to v. Hadamard matrices are type-II matrices.

If W is type II, then so are W^T and $W^{(-)}$. If D is diagonal and invertible, then DW and WD are both type II; if P is a permutation matrix then PW and WP are type-II. If W_1 and W_2 are type-II matrices, so is their Kronecker product $W_1 \otimes W_2$.

We say that a complex matrix is *flat* if all its entries have the same absolute value. Hadamard matrices are flat.

23.4.1 Lemma. *If W is a square complex matrix, then any two of the following imply the third:*

(a) W is type II.

(b) A non-zero scalar multiple of W is unitary.

(c) W is flat.

Proof. Exercise. □

Suppose $x_1, \ldots, x_n$ is an equiangular set of n lines in $\mathbb{C}^d$ with squared cosine γ and the relative bound is tight. Let G be the Gram matrix of a set of unit vectors spanning these lines and set

$$S = \frac{1}{\sqrt{\gamma}}(G - I).$$

From the previous section, the eigenvalues of S are

$$\tau := -\frac{1}{\sqrt{\gamma}}, \quad \theta := \frac{n-d}{d\sqrt{\gamma}}$$

and so $(S - \tau I)(S - \theta I) = 0$ and hence

$$S^2 = (\theta + \tau)S - \theta\tau I = \frac{n - 2d}{d\sqrt{\gamma}} S + (n - 1)I.$$

23.4.2 Lemma. *Suppose S is the Seidel matrix of an equiangular set of n lines in $\mathbb{C}^d$ with squared cosine γ. If the relative bound is tight and*

$$\lambda + \lambda^{-1} + \frac{n - 2d}{d\sqrt{\gamma}} = 0,$$

then $\lambda I + S$ is a type-II matrix.

Proof. We note that

$$S = S^* = S^{(-)T}$$

and therefore

$$\begin{aligned}(\lambda I + S)(\lambda I + S)^{(-)T} &= (\lambda I + S)(\lambda^{-1} I + S^{(-)T}) \\ &= (\lambda I + S)(\lambda^{-1} I + S) \\ &= I + (\lambda + \lambda^{-1})S + S^2 \\ &= I(1 - \theta\tau) + (\lambda + \lambda^{-1} + \theta + \tau)S.\end{aligned}$$

The lemma follows immediately. □

Let us call a matrix *d-flat* if its off-diagonal entries all have the same absolute value and its diagonal entries all have the same absolute value. We say that a d-flat matrix is*normalized* if it off-diagonal entries all have absolute value 1, and its diagonal entries are real.

23.4.3 Lemma. *Suppose W is a normalized d-flat type-II matrix. If $W_{i,i} = \delta \neq 1$ and $S := W - \delta I$, then S is the Gram matrix of a set of equiangular lines realizing the relative bound.*

Proof. Assume W is $v \times v$. We see that $S^{(-)T} = S^*$ and so

$$W^{(-)T} = \delta^{-1} I + S^*.$$

Therefore

$$vI = WW^{(-)T} = I + \delta S^* + \delta^{-1} S + SS^* \tag{23.4.1}$$

and on taking the conjugate-transpose of this, we get

$$vI = I + \delta S + \delta^{-1} S^* + SS^*.$$

Comparing this with (23.4.1) yields that

$$(\delta - \delta^{-1})S = (\delta - \delta^{-1})S^*.$$

Therefore S is Hermitian and so (23.4.1) implies that

$$S^2 + (\delta + \delta^{-1})S - (v-1)I = 0.$$

If τ is the least eigenvalue of S, then

$$I + \frac{1}{\tau} S$$

is positive semidefinite with one positive eigenvalue. □

23.5 The Unitary Group

Let V be an inner product space. A linear operator M on V is *orthogonal* if

$$\langle Mu, Mv \rangle = \langle u, v \rangle$$

for all u and v in V. If the inner product is complex we often call an orthogonal operator *unitary*. The matrix that represents an orthogonal operator relative to an orthogonal basis is also said to be orthogonal. The *adjoint* M^* of M is the operator defined by the condition

$$\langle U, Mv \rangle = \langle M^* u, v \rangle.$$

Thus M is unitary if and only if $M^* = M^{-1}$.

An orthogonal operator clearly preserves the length of a vector.

23.5.1 Lemma. *A linear operator preserves length if and only if it is orthogonal.*

Proof. We assume the inner product is complex, as this case is slightly trickier and it implies the real case.

If M is a unitary linear operator on V and $x, y \in V$, then

$$\begin{aligned}\langle x+y, x+y\rangle &= \langle Mx+My, Mx+My\rangle \\ &= \langle Mx, Mx\rangle + \langle My, My\rangle + \langle Mx, My\rangle + \langle My, Mx\rangle \\ &= \langle x, x\rangle + \langle y, y\rangle + \langle Mx, My\rangle + \langle My, Mx\rangle\end{aligned}$$

Therefore

$$\langle x, y\rangle + \langle y, x\rangle = \langle Mx, My\rangle + \langle My, Mx\rangle$$

Setting ix in place of x in this identity yields

$$-i\langle x, y\rangle + i\langle y, x\rangle = -i\langle Mx, My\rangle + i\langle My, Mx\rangle$$

and therefore

$$\langle x, y\rangle = \langle Mx, My\rangle$$

for all x and y. □

We construct a useful class of unitary mappings. Suppose $\varphi \in V^* \setminus 0$ and $y \in V \setminus 0$. If we define the mapping τ by

$$\tau(x) := x + \varphi(x)y$$

then τ is linear and fixes each vector in $\ker \varphi$. Then

$$\langle \tau(x), \tau(x)\rangle = \langle x, x\rangle + \varphi(x)\langle x, y\rangle + \overline{\varphi(x)}\langle y, x\rangle + \varphi(x)\overline{\varphi(x)}\langle y, y\rangle \quad (23.5.1)$$

and if τ preserves length, then $\varphi(x) = 0$ whenever $\langle y, x\rangle = 0$. Thus the kernel of the linear map

$$x \mapsto \langle y, x\rangle$$

is contained in $\ker(\varphi)$. Since both kernels have codimension one in V, they are equal and consequently there is a non-zero scalar λ such that

$$\varphi(x) = \lambda\langle y, x\rangle$$

for all x.

If we substitute $\phi(x) = -\lambda\langle y, x\rangle/\langle y, y\rangle$ in (23.5.1) then

$$\langle \tau(x), \tau(x)\rangle = \langle x, x\rangle - (\lambda + \overline{\lambda} - \lambda\overline{\lambda})\frac{\langle x, y\rangle\langle y, x\rangle}{\langle y, y\rangle}$$

and therefore τ preserves length if and only if

$$\lambda + \overline{\lambda} - \lambda\overline{\lambda} = 0,$$

which happens if and only if $\|1 - \lambda\| = 1$.

23.5.2 Corollary. *The map $\tau : V \to V$ given by*

$$\tau(x) = x - 2\frac{\langle y, x\rangle}{\langle y, y\rangle}y$$

is unitary. □

We note that

$$\tau(y) = y - 2y = -y$$

and from this it follows that $\tau^2 = 1$. The map τ is therefore called a *complex reflection*.

23.6 A Special Group

A diagonal matrix is unitary if its diagonal entries have absolute value 1. If D is a diagonal matrix and P a permutation matrix of the same order, then $PDP^T = D'$ is diagonal and so $PD = D'P$. Define two $d \times d$ matrices X and Y as follows. Let $e_0, \ldots, e_{d-1}$ be the standard basis for $\mathbb{C}^d$, with the understanding that the indices $0, \ldots, d-1$ are integers modulo d, and set $\theta = \exp(2\pi i/d)$. Then

$$Xe_i = e_{i+1}, \qquad Ye_i = \theta^i e_i.$$

Thus X is a permutation matrix and Y is diagonal. The *Weyl-Heisenberg* is the group generated by X, Y and θI. (A physicist would call it a *generalized Pauli group*.)

We investigate some of the properties of this group, which we denote by G. We calculate

$$XYe_i = \theta^i e_{i+1}, \qquad YXe_i = \theta^{i+1} e_{i+1},$$

and thus

$$YX = \theta XY,$$

in particular X and Y do not commute. It also follows from this relation that each element of our group can be written in the form

$$\theta^r X^s Y^t$$

where $0 \leq r, s, t \leq d-1$. This shows that $|G| \leq d^3$. You should prove that equality holds.

The subgroup if G consisting of the elements $\theta^r I$ is central with order d. The quotient G/D is abelian and is isomorphic to $\mathbb{Z}_d^2$.

23.6.1 Lemma. *The group G acts irreducibly on $\mathbb{C}^d$.*

Proof. Suppose U is a non-zero subspace that is fixed by G, and let u be a non-zero vector in it. Then U contains $X^r u$ for all r, and so we may assume without loss that $u_1 \neq 0$. Then the vector

$$\sum_r Y^r u = ce_1,$$

for some scalar c. Hence U contains $_1$, and since $e_r = X^r e_1$, we conclude that U contains a basis for $\mathbb{C}^d$. Therefore $U = \mathbb{C}^d$, and so we have shown that no proper subspace of $\mathbb{C}^d$ is G-invariant. □

This lemma has important consequences. First, it implies that the only matrices that commute with all elements of G are the scalar matrices. Second it implies that the subspace of $\mathrm{Mat}_{d\times d}(\mathbb{C})$ spanned by the elements of G has dimension d, and thus this subspace is $\mathrm{Mat}_{d\times d}(\mathbb{C})$. (The second fact, due to Burnside, implies the first, due to Schur.)

The Weyl-Heisenberg group can be used to construct sets of d^2 equiangular lines in $\mathbb{C}^d$. The idea is to choose a non-zero vector f in $\mathbb{C}^d$, and consider the lines spanned by the images of f under the action of the d^3 elements of the group. This will produce at most d^2 lines, and in certain cases the result is a set of d^2 equiangular lines. To be more specific, if $d = 2$ we make take f to be one of the two vectors

$$\frac{1}{\sqrt{6}} \begin{pmatrix} \pm\sqrt{3 \pm \sqrt{3}} \\ e^{i\pi/4}\sqrt{3 \mp \sqrt{3}} \end{pmatrix}$$

When $d = 3$ we make take f to be

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

There is no algorithm for finding f, but examples have been constructed in dimensions $2-7$ and 18. An example is also known when $d = 8$, although it uses a different group. The physicists call f a *fiducial vector*.

Wootters http://arxiv.org/pdf/quant-ph/0406032

Renes, Blume-Kohout, Scott, Caves http://arxiv.org/abs/quant-ph/0310075

Flammia http://arxiv.org/pdf/quant-ph/0605050

Chapter 24

Spherical Designs

24.1 Orthogonal Polynomials

We work with real inner products on the vector space of polynomials $\mathbb{R}[t]$, or the subspace of real polynomials with degree at most n. We assume that this inner product satisfies

$$\langle p, tq\rangle = \langle tp, q\rangle.$$

Multiplication by t is a linear endomorphism of $\mathbb{R}[t]$, the given condition asserts that this endomorphism is self-adjoint relative to the given inner product. We will also assume that if f is a non-zero polynomial and $f(t) \geq 0$ for all t, then

$$\langle 1, f\rangle > 0.$$

Examples. If $w(t)$ is a non-negative real function such that for non-negative integers m,

$$\int t^{2m} w(t)\, dt < \infty$$

then we may take

$$\langle p, q\rangle := \int p(t)q(t)w(t)\, dt.$$

As particular cases, we might have

$$\langle p, q\rangle = \int_{-1}^{1} (1-t^2)^{m/2}\, dt$$

or

$$\langle p, q\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} p(t)q(t)\exp(-t^2/2)\, dt.$$

A sequence of polynomials p_n $(n \geq 0)$ is a sequence of *orthogonal polynomials* if

(a) $\deg(p_n) = n$ for all n.

(b) If $i \neq j$, then $\langle p_i, p_j \rangle = 0$.

Any such sequence can be constructed by using Gram-Schmidt to obtain an orthogonal basis for $\mathbb{R}[t]$ from the basis

$$1,\ t,\ t^2, \ldots$$

Note that this sequence is not completely determined; if we multiply each term in a sequence of orthogonal polynomials by a non-zero scalar, the resulting sequence is still a sequence of orthogonal polynomials. We can eliminate this ambiguity by normalizing the polynomials. There are three common ways to do this. We could take the polynomials to be monic, we might arrange that they be orthonormal, or we could choose a scalar a which is not a zero of any of the polynomials and then assume that $p_n(a) = 1$ for all n.

24.2 A Three-Term Recurrence

24.2.1 Theorem. *If $(p_n)_{n\geq 0}$ is a sequence of orthogonal polynomials, then there are scalars a_n, b_n and c_n such that for each n,*

$$tp_n(t) = b_n p_{n-1}(t) + a_n p_n(t) + c_n p_{n+1}(t).$$

Proof. We first observe that p_n is orthogonal to all polynomials of degree less then n. So if $r < n - 1$, then

$$0 = \langle p_n, tp_r \rangle = \langle tp_n, p_r \rangle$$

Similarly if $r > n + 1$, then $\langle tp_n, p_r \rangle = 0$. Therefore tp_n must be a linear combination of p_{n-1}, p_n and p_{n+1}. □

If we normalize our polynomials so that p_n is monic for all n, then $c_n = 1$ for all n. Further, the coefficient of p_n in the expansion of tp_{n-1} as a linear combination of orthogonal polynomials is 1. Hence

$$\langle tp_n, p_{n-1} \rangle = \langle p_n, tp_{n-1} \rangle = \langle p_n, p_n \rangle > 0$$

and consequently

$$b_n = \frac{\langle p_n, p_n \rangle}{\langle p_{n-1}, p_{n-1} \rangle}.$$

We also have

$$a_n = \frac{\langle p_n, tp_n \rangle}{\langle p_n, p_n \rangle}.$$

24.2.2 Lemma. *Suppose p_n is a member of a sequence of orthogonal polynomials and $q \mid p_n$. If $q \geq 0$, then q is constant.*

Proof. Assume $p = qh$. Then

$$\langle p, h \rangle = \langle 1, ph \rangle = \langle 1, qh^2 \rangle$$

and since $qh^2 \geq 0$, it follows that $\langle p, h \rangle > 0$. On the other hand if q is not constant then $\delta(h) < n$ and $\langle p, h \rangle = 0$. □

24.2.3 Theorem. *If p_n is a member of a sequence of orthogonal polynomials, then its zeros are real and simple.*

Proof. If the zeros of p_n are not all real, then since its coefficients are real it must have a complex conjugate pair of zeros, θ and $\bar{\theta}$ say. So

$$q(t) := (t - \theta)(t - \bar{\theta})$$

is a real factor of p_n that is non-negative for all t. By the previous lemma, this is impossible.

If θ is a zero of p_n that is not simple, then

$$q(t) = (t - \theta)^2$$

is a real factor of p_n that is non-negative for all t. □

Using the three-term recurrence, we could prove that the zeros of p_{n-1} are real and interlace the zero of p_n.

24.3 The Unit Sphere

24.3.1 Lemma. $\mathrm{Pol}(\Omega, r) = \mathrm{Hom}(r) \oplus (x_1^2 + \cdots + x_d^2)\,\mathrm{Hom}(r-1)$.

24.3.2 Corollary.

$$\dim \mathrm{Pol}(\Omega, r) = \binom{d+r-1}{r} + \binom{d+r-2}{r-1}.$$

Let Ω denote the unit sphere in $\mathbb{R}^d$. If f and g are functions on Ω, we define

$$\langle f, g\rangle = \int fg\, d\mu$$

where μ denotes the usual measure on Ω. Note that

$$\langle 1, f\rangle$$

is the average value of f on Ω.

If $a \in \Omega$ and $p \in \mathbb{R}[t]$, we define the function p_a by

$$p_a(x) := p(a^T x).$$

We say that p_a is a *zonal polynomial* relative to a.

If we fix a then the inner product on $\mathrm{Pol}(\Omega,)$ gives an inner product on $\mathbb{R}[t]$. The corresponding family of orthogonal polynomials is known as the *Gegenbauer polynomials*. Here the inner product is given by

$$\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,(1-t^2)^{(d-3)/2} dt.$$

If g_i denotes the i-th Gegenbauer polynomial, we use $g_{a,i}$ to denote $(g_i)_a$. We will refer to $g_{a,i}$ as a Gegenbauer polynomial. Note that if $g_i(1) = 0$, then $1 - t$ is a non-negative factor of $g_i(t)$. It follows that $g_i(1) \neq 0$ and hence we may choose our normalization so that

$$\langle g_{a,i}, g_{a,i}\rangle = g_{a,i}(a).$$

The polynomials $g_{a,i}$ are known as *zonal orthogonal polynomials* on the sphere.

We note that

$$\begin{aligned}
g_0(t) &= 1\\
g_1(t) &= dt\\
2g_2(t) &= (d+2)(dt^2-1)\\
6g_3(t) &= d(d+4)[(d+2)t^3 - 3t]\\
24g_4(t) &= d(d+2)(d+8)[(d+2)(d+4)t^4 - 6(d+2)t^2 + 3]
\end{aligned}$$

24.4 Two Bounds

Suppose $\Phi \subseteq \Omega$. The *degree* of Φ is the size of the set

$$\{x^T y : x, y \in \Phi,\ x \neq y\}.$$

We call the set the *degree set* of Φ. If $f \in \mathrm{Pol}(\Omega,)$, then

$$\langle 1, f\rangle_\Phi := \frac{1}{|\Phi|}\sum_{x\in\Phi} f(x).$$

We say that Φ has *strength* at least r if

$$\langle 1, f\rangle = \langle 1, f\rangle_\Phi$$

for all f in $\mathrm{Pol}(\Omega, r)$.

24.4.1 Theorem. *If $\Phi \subseteq \Omega$ and $\deg(\Phi) = s$, then*

$$|\Phi| \leq \dim \mathrm{Pol}(\Omega, r).$$

Proof. If D is the degree set of Φ, we define

$$\varphi(t) := \prod_{\lambda\in D} \frac{t-\lambda}{1-\lambda}.$$

Then $\varphi_a(a) = 1$ and $\varphi_a(b) = 0$ if $b \in \Phi \setminus a$. Therefore the restrictions to Φ of the functions φ_a for a in Φ are linearly independent, and so the functions φ_a are linearly independent. □

24.4.2 Theorem. *If Φ is a subset of Ω with strength r, then*

$$|\Phi| \geq \dim(\mathrm{Pol}(\Omega, \left\lfloor \frac{t}{2} \right\rfloor)).$$

Proof. Let $h_1, \ldots, h_n$ be an orthonormal basis for $\mathrm{Pol}(\Omega, \left\lfloor \frac{t}{2} \right\rfloor))$. Then $h_i h_j$ has degree at most t and consequently

$$\langle h_i, h_j\rangle = \langle h_i, h_j\rangle_\Phi.$$

Therefore the restrictions to Φ of the functions h_i are orthogonal, and therefore these restrictions are linearly independent elements of the space of real functions on Φ, whose dimension is $|\Phi|$. □

24.5 Harmonic Polynomials

We define $\mathrm{Harm}(r)$ to be the orthogonal complement to $\mathrm{Pol}(\Omega, r-1)$ in $\mathrm{Pol}(\Omega, r)$. The elements of $\mathrm{Harm}(r)$ are *harmonic polynomials* of degree r.

If $f \in \mathrm{Pol}(\Omega,)$, let $P_a(f)$ denote the orthogonal projection of f onto the space of zonal polynomials relative to a. If γ lies in the orthogonal group $O(d)$ we define f^γ by

$$f^\gamma(x) := f(\gamma x).$$

We make two claims:

(a) $P_a(f)$ is equal to the average of the functions f^γ, where γ runs over the subgroup of $O(d)$ that leaves a fixed.

(b) $\deg(P_a(f)) \le \deg(f)$.

24.5.1 Theorem. *If $\rho_{a,r} := \sum_{j \le r} g_{a,j}$ and $\deg(f) \le r$, then $\langle f, \rho_{a,r}\rangle = f(a)$.*

Proof. We have

$$\langle P_a(f), g_{a,i}\rangle = \langle f, g_{a,i}\rangle$$

and therefore

$$P_a(f) = \sum_{i \le r} \frac{\langle f, g_{a,i}\rangle}{\langle g_{a,i}, g_{a,i}\rangle} g_{a,i}.$$

Since $\langle g_{a,i}, g_{a,i}\rangle = g_{a,i}$, this implies that

$$P_a(f)(a) = \sum_{i \le r} \langle f, g_{a,i}\rangle = \langle f, \rho_r\rangle.$$

To complete the proof, note that $P_a(f)(a) = f(a)$. □

The next result is known as the *addition rule*, and plays a very important role.

24.5.2 Corollary.

$$\langle g_{a,i}, g_{b,j}\rangle = \delta_{i,j} g_{a,i}(b).$$

Proof. If $j < i$, then $\deg(P_a(g_{b,i})) < i$ and therefore

$$\langle g_{a,i}, g_{b,j}\rangle = \langle g_{a,i}, P_a(g_{b,j})\rangle = 0.$$

If $i = j$, then

$$\langle g_{a,i}, g_{b,i}\rangle = \langle \rho_{a,i}, g_{b,i}\rangle = g_{b,i}(a) = g_{a,i}(b).$$ □

24.5.3 Theorem. *If $p_1, \ldots, p_n$ is an orthonormal basis for* $\mathrm{Harm}(r)$*, then*

$$g_{a,r}(x) = \sum_{i=1}^{n} h_i(a) h_i(x).$$

Proof. Since $\langle h_i, h_i \rangle = 1$, we have

$$g_{a,r} = \sum_{i=1}^{n} \langle g_{a,r}, p_i \rangle p_i = \sum_{i=1}^{n} \langle h_{a,r}, p_i \rangle p_i = \sum_{i=1}^{n} h_i(a) h_i. \qquad \square$$

24.6 Linear Programming

24.6.1 Lemma. *If Φ is a finite subset of Ω, then $\sum_{a,b \in \Phi} g_{a,r}(b) \geq 0$.*

Proof. We have

$$\sum_{a,b \in \Phi} g_{a,r}(b) = \sum_{a,b \in \Phi} \sum_{i} h_i(a) h_i(b) = \sum_{i} \left(\sum_{a \in \Phi} h_i(a) \right)^2 \geq 0. \qquad \square$$

Essentially the same argument shows that the kernel $g_{a,r}(b)$ is positive semidefinite.

24.6.2 Theorem. *Suppose Φ is a subset of Ω with degree s and let $F(t)$ be a polynomial such that*

(a) $\deg F \leq s$.

(b) $F(1) = 1$.

(c) If $a, b \in \Phi$ and $a \neq b$, then $F_a(b) \leq 0$.

(d) $\langle F, g_{a,i} \rangle \geq 0$.

Then

$$|\Phi| \leq \frac{F(1)}{\langle 1, F_a \rangle}.$$

Proof. Assume

$$F_a = \sum_{i=0}^{s} f_i g_{a,i}.$$

We have

$$\langle 1, F_a\rangle_\Phi = \frac{1}{|\Phi|}\Big(1 + \sum_{b\in\Phi\setminus a} F_a(b)\Big) \le \frac{1}{|\Phi|}$$

and therefore

$$\begin{aligned}\frac{1}{|\Phi|} &\ge \frac{1}{|\Phi|}\sum_{a\in\Phi}\langle 1, F_a\rangle_\Phi \\ &= \frac{1}{|\Phi|}\sum_{a\in\Phi}\sum_{i=0}^{s} f_i\langle 1, g_{a,i}\rangle_\Phi \\ &= \sum_{i=0}^{s}\frac{f_i}{|\Phi|^2}\Big(\sum_{a,b\in\Phi} g_{a,i}(b)\Big) \\ &\ge f_0.\end{aligned}$$

□

By way of example, suppose

$$F(t) = (t\alpha)(t-\beta).$$

Then

$$f_0 = \alpha\beta + \frac{1}{d}, \quad f_1 = -\frac{\alpha+\beta}{d}, \quad f_2 = \frac{2}{d(d+1)}$$

and consequently if $-1 \le \alpha, \beta \le 1$ and

$$\alpha + \beta \le 0, \quad \alpha\beta \ge \frac{1}{d},$$

then a subset of Ω with degree set contained in $\{\alpha, \beta\}$ has size at most

$$\frac{d(1-\alpha)(1-\beta)}{1+d\alpha\beta}.$$

Chapter 25

Frames

25.1 Isoclinic Subspaces

Let U and V be two k-dimensional subspace of an inner product space W, and let P and Q be the corresponding orthogonal projections. Then P maps the unit sphere in V to an ellipsoid in U. The shape of this ellipsoid is determined by the extreme points of the function

$$\|Pv\|^2 = v^*P^*Pv = v^*Pv,$$

where v runs over the unit vectors in V. We say that V is *isoclinic* to U is there is a constant λ such that

$$v^*Pv = \lambda v^*v.$$

If V is isoclinic to U with parameter λ, then

$$x^*Q^*PQx = \lambda x^*Q^*QX = \lambda x^*Qx$$

for all x in w. Hence we see see that U and V are isoclinic with parameter λ if and only if

$$QPQ = \lambda Q.$$

Thus we have translated a geometric condition into a linear algebraic one. Our next result shows that is a symmetric relation.

25.1.1 Lemma. *The subspace U is isoclinic to V if and only if V is isoclinic to U.*

Proof. Let R be a matrix whose columns form an orthonormal basis for U, and let S be a matrix whose columns form an orthonormal basis for V. Then

$$RR^* = P, \quad SS^* = Q$$

and

$$QPQ = SS^*RR^*SS^* = S(S^*RR^*S)S^*.$$

If $QPQ = \lambda Q$, then it follows that

$$\lambda SS^* = S(S^*RR^*S)S^*$$

and therefore

$$\lambda I = S^*S(S^*RR^*S)S^*S = S^*RR^*S.$$

Hence $R^*SS^*R = \lambda I$ and so

$$\lambda P = \lambda RR^* = R(R^*SS^*R)R^* = PQP. \qquad \square$$

Note that $\mathrm{tr}(PQP) = \mathrm{tr}(QPQ)$, and so if $\mathrm{rk}(P) = \mathrm{rk}(Q)$ and $QPQ = \lambda P$, then $PQP = \lambda P$. A consequence of the proof is that U and V are isoclinic if and only the matrix $\lambda^{-1}R^*S$ is orthogonal.

As exercises, prove that if P and Q are projections then $(P-Q)^2$ commutes with P and Q. Also if U and V are isoclinic with parameter λ, then

$$(P-Q)^3 = (1-\lambda)(P-Q).$$

This implies that the eigenvalues of $P-Q$ are

$$0, \ \pm\sqrt{1-\lambda};$$

since $\mathrm{tr}(P-Q) = 0$, the non-zero eigenvalues have equal multiplicity. Unfortunately I have no idea what to do with this information :-(

25.2 Matrices

We investigate sets of pairwise isoclinic k-subspaces in $\mathbb{R}^n$. Let U be the column space of the matrix

$$R = \begin{pmatrix} I_k \\ 0 \end{pmatrix}.$$

Suppose S is the $n \times k$ matrix

$$S = \begin{pmatrix} Y \\ Z \end{pmatrix}$$

where $S^*S = I_k$. Then the column spaces of R and S are λ-isoclinic if and only if

$$\lambda I = S^*RR^*S = Y^*Y.$$

Since

$$I = S^*S = Y^*Y + Z^*Z$$

we then have $Z^*Z = (1-\lambda)I$. If

$$T = \begin{pmatrix} \lambda^{1/2} I \\ \lambda^{-1/2} ZY^* \end{pmatrix}$$

then $T = \lambda^{-1/2} SY^*$, so $\mathrm{col}(T) = \mathrm{col}(S)$ and $T^*T = I$.

25.2.1 Lemma. *If V is λ-isoclinic to the column space of*

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

then V is the column space of a matrix

$$\begin{pmatrix} \lambda^{1/2} I_k \\ \lambda^{-1/2} Z \end{pmatrix}$$

where $Z^*Z = (1-\lambda)I$.

Now suppose

$$a^2 I + A^*A = b^2 I + B^*B = I;$$

then the column spaces of the matrices

$$R = \begin{pmatrix} aI \\ A \end{pmatrix}, \quad S = \begin{pmatrix} bI \\ B \end{pmatrix}$$

are isoclinic if and only if R^*S is a scalar multiple of an orthogonal matrix. We have

$$R^*S = abI + A^*B$$

and so our spaces are ν-isoclinic if and only if

$$(abI + A^*B)(abI + B^*A) = \nu I.$$

Equivalently, $\nu^{-1/2}(I + A^*B)$ must be unitary.

25.3 Equiangular Subspaces

Suppose that $P_1, \ldots, P_m$ are projections onto e-dimensional subspaces of d-dimensional vector space. We say that they are *equiangular* if there is a scalar α^2 such that

$$\operatorname{tr}(P_i P_j) = \alpha^2$$

whenever $i \neq j$. We note that

$$\operatorname{tr}(P - Q)^2 = 2e - 2\operatorname{tr}(PQ)$$

where $\operatorname{tr}(P-Q)^2$ is the Euclidean distance between the matrices P and Q. So we could have used "equidistant" in place of "equiangular".

25.3.1 Lemma. *An equiangular set of projections is linearly independent.*

Proof. Suppose we have scalars $c_1, \ldots, c_m$ such that

$$0 = \sum_i c_i P_i.$$

Then

$$0 = \sum_i \operatorname{tr}(P_r P_i) = c_r e + \alpha^2 \sum_{i \neq r} c_i = e(c_r - \alpha^2) + \alpha_2 \sum_i c_i.$$

From this we deduce that c_r is independent of r and hence that $c_r = 0$ for all r. □

The projections P_i are Hermitian and so, if we work over $\mathbb{C}$, they lie in a real vector space of dimension d^2. Over $\mathbb{R}$ they lie in a space of dimension $d(d+1)/2$. These upper bounds are known as the *absolute bounds*. The bound supplied by the following theorem is the *relative bound*.

25.3.2 Theorem. *If the projections $P_1, \ldots, P_m$ are equiangular with angle α^2 and $d\alpha^2 \leq e$, then*

$$m \leq \frac{d(e - \alpha^2)}{e^2 - d\alpha^2},$$

equality holds if and only if

$$\sum_i P_i = \frac{me}{d} I.$$

Proof. We set

$$S := \sum_i \Big(P_i - \frac{e}{d} I \Big)$$

Then $S = S^*$ and therefore $\operatorname{tr}(S^2) \geq 0$, which yields

$$\begin{aligned} 0 &\leq \sum_i \operatorname{tr}\Big(P_i - \frac{e}{d} I \Big)^2 + \sum_{i \neq j} \operatorname{tr}\Big[\Big(P_i - \frac{e}{d} I \Big) \Big(P_j - \frac{e}{d} I \Big) \Big] \\ &= m\Big(e - \frac{e^2}{d} \Big) + m(m-1)\Big(\alpha^2 - \frac{e^2}{d} \Big). \end{aligned}$$

Our bound follows from this. If equality holds that $\operatorname{tr}(S^2) = 0$ and therefore $S = 0$. □

If P and Q are projections onto isoclinic spaces with parameter λ, then

$$\lambda e = \operatorname{tr}(\lambda P) = \operatorname{tr}(PQP) = \operatorname{tr}(PQ) = \alpha^2.$$

Thus $\lambda = \alpha^2/e$ and our expression for m becomes

$$m = \frac{d(1-\lambda)}{e - d\lambda}.$$

This bound (for equi-isoclinic subspaces) is due to Lemmens and Seidel. They also note that the absolute bound cannot be tight if $e > 1$, because the projections P_i lie in the subspace of mappings Q such that $P_1 Q P_1$ is a scalar multiple of Q and this has codimension $e(e+1)/2$.

A set $P_1, \ldots, P_m$ of projections with rank e such that

$$\sum P_i = \frac{me}{d}$$

is known as a *tight fusion frame*. If $e = 1$, it is a *tight frame*.

If R_i is a matrix whose columns form an orthonormal basis for $\operatorname{im}(P_i)$, then

$$P_i = R_i R_i^*.$$

So if $\sum_i P_i = (me/d)I$, then

$$\frac{me}{d} I = \sum_i R_i R_i^*.$$

If $\mathcal{R}$ denotes the $d \times me$ matrix

$$\begin{pmatrix} R_1 & \ldots & R_m \end{pmatrix}$$

then

$$\mathcal{R}\mathcal{R}^* = \sum_i R_i R_i^* = \frac{me}{d} I$$

and accordingly $\mathcal{R}^*\mathcal{R}$ is a scalar multiple of a projection of order $me \times me$. (It has a block decomposition where the ij-block is $R_i^* R_j$; this block is a scalar multiple of an orthogonal matrix.)

25.4 Tight Frames

Let V be an inner product space. A sequence of vectors $x_1, \ldots, x_m$ is a *frame* if there are positive reals A and B such that, for any vector z in V we have:

$$A\|z\|^2 \leq \sum_i |\langle x_i, z\rangle|^2 \leq B\|z\|^2.$$

If we define

$$S := \sum x_i x_i^*$$

then the inner term in the inequalities above is equal to $z^* S z$ and the best choice for A and B will the be the least and largest eigenvalues of S. The matrix S is called the *frame operator*. It is invertible if and only if the vectors x_i span V. A frame is *tight* if $A = B$ or, equivalently, if S is a scalar matrix. A frame is *uniform* if all vectors in it have the same norm.

As an exercise, show that any frame can be extended to a tight frame by adding at most $\dim(V) - 1$ vectors.

25.4.1 Lemma. *If the size of a set of equiangular lines meets the relative bound, then the corresponding projections form a tight frame.* □

25.4.2 Theorem. *A set of unit vectors forms a tight uniform frame if and only if it is a spherical 2-design.*

Chapter 26

276

In this chapter we consider equiangular sets of lines in real space that meet the absolute bound.

26.1 Cocliques

A *coclique* in a graphs is a set of vertices such that no two are adjacent. Cocliques are also known as independent sets. Here our concern is with cocliques in regular two graphs. The maximim size of a coclique in X is denoted by $\alpha(X)$.

Suppose X is the graph of a set of equiangular lines in $\mathbb{R}^d$, with squared cosine γ. The vertices of X are unit vectors, and the unit vectors $x_1, \ldots, x_s$ form a coclique if and only if

$$\langle x_i, x_j \rangle = -\sqrt{\gamma}$$

when $i \neq j$. Hence the submatrix of the Gram matrix of the vertices of X is equal to

$$I - \sqrt{\gamma}(J - I) = (1 + \sqrt{\gamma})I - \sqrt{\gamma}J$$

Since this is a principal submatrix of a Gram matrix, it must be positive semidefinite and hence its eigenvalues are non-negative. Now the eigenvalues of this matrix are

$$1 + \sqrt{\gamma}, \quad 1 - (s-1)\sqrt{\gamma}$$

and we have the following:

26.1.1 Lemma. *If X is the graph of a set of equiangular lines with square cosine γ, then*

$$\alpha(X) \leq 1 + \frac{1}{\sqrt{\gamma}}.$$

We can deduce more when equality holds. In this case the row sums of $I - \sqrt{\gamma}(J - I)$ are zero, and so if this matrix is the Gram matrix of the unit vectors $x_1, \ldots, x_r$, then the sum of these vectors must be zero. If y is a vertex of X then

$$\sum_{i=1}^{s} \langle y, x_i \rangle = \left\langle y, \sum_{i=1}^{s} x_i \right\rangle = 0$$

and if neither y nor $-y$ lies in S, then $\langle y, x_i \rangle = \pm\sqrt{\gamma}$ and so if y has exactly r neighbors in S, then

$$0 = r\sqrt{\gamma} + (s - r)(-\sqrt{\gamma}) = 2r\sqrt{\gamma} - s\sqrt{\gamma}$$

and hence $r = s/2$. It follows that each vertex not in S or $-S$ is adjacent to exactly half the vertices in S.

Chapter 27

Mutually Unbiased Bases

Suppose $x_1, \ldots, x_d$ and $y_1, \ldots, y_d$ are two orthonormal bases of some inner product space. We say these two bases are *mutually unbiased* if there is a real scalar γ such that, for all i and j,

$$|\langle x_i, y_j \rangle|^2 = \gamma.$$

We will call γ the squared cosine of the set of bases. A set of bases is mutually unbiased if each pair from it is unbiased. The columns of the two matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

form a mutually unbiased pair of bases in $\mathbb{R}^2$.

27.1 Basics

Suppose $x_1, \ldots, x_d$ and $y_1, \ldots, y_d$ are a pair of mutually unbiased bases. If

$$y_j = \sum_i c_i x_i$$

then

$$1 = \|y_i\|^2 = \sum_{i=1}^{d} |c_i|^2 = d\gamma.$$

Therefore

$$\gamma = \frac{1}{d}.$$

The columns of the unitary matrices M and N form a mutually unbiased pair of bases if and only if the columns of I and $M^{-1}N$ do. Thus any set of r mutually unbiased bases can be specified by a set of r unitary matrices, one of which is the identity.

27.1.1 Lemma. *If M is unitary and I and M are unbiased, then A is flat.* □

Recall that unitary matrix is a type-II matrix if and only if it is flat. Thus each flat unitary matrix determines a mutually unbiased pair of bases.

27.2 Bounds

27.2.1 Theorem. *A set of mutually unbiased bases in $\mathbb{C}^d$ contains at most $d+1$ bases; in $\mathbb{R}^d$ we have at most $\frac{d}{2}+1$ bases.*

Proof. Suppose we have vectors $x_{i,j}$ where $1 \leq i \leq m$ and for each i, the vectors $x_{i,1}, \ldots, x_{i,d}$ form an orthonormal basis. Assume further that these bases are mutually unbiased. Let $X_{i,j}$ denote the projection corresponding to $x_{i,j}$ and let G be the Gram matrix of the projections. Then G has the form

$$I_{md} + \gamma((J_m - I_m) \otimes J_d).$$

We determine the rank of G. Its eigenvalues are

$$\gamma(m-1)d + 1,\ 1 - \gamma d,\ 1$$

with respective multiplicities 1, $m-1$ and $md-m$. As $d\gamma = 1$, we see that $\mathrm{rk}(G) = md - d + 1$. Hence the projections $X_{i,j}$ span a subspace of the space of Hermitian matrices with dimension $md - m + 1$ and so, in the complex case,

$$md - m + 1 \leq d^2,$$

from which it follows that $m \leq d+1$. In the real case we get $m \leq (d+2)/2$. □

27.3 MUB's

If $x_1, \ldots, x_d$ and $y_1, \ldots, y_d$ are two orthonormal bases in $\mathbb{C}^d$, we say that they are *unbiased* if there is a constant γ such that for all i and j,

$$\langle x_i, y_j \rangle \langle y_j, x_i \rangle = \gamma.$$

In other words, the angle between any two lines spanned by vectors in different bases is the same. A set of orthonormal bases is *mutually unbiased* if each pair of bases in it is unbiased. If U and V are $d \times d$ unitary matrices then their columns provide a pair of orthonormal bases, and these bases are unbiased if and only if the matrix U^*V is flat. Note that U^*V is itself unitary, and that its columns and the standard basis of $\mathbb{C}^d$ are unbiased.

The two bases

$$\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}; \quad \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$$

are mutually unbiased.

The angle between lines corresponding to vectors from distinct orthogonal bases is determined by d. To see this, suppose $x_1, \ldots, x_d$ and $y_1, \ldots, y_d$ are orthogonal and unbiased with $|\langle x_i, y_j\rangle|^2 = \gamma$. Then since $x_1, \ldots, x_d$ is an orthonormal basis

$$y_1 = \sum \langle x_i, y_1\rangle x_i$$

and

$$\langle y_1, y_1\rangle = \sum_i |\langle x_i, y_1\rangle|^2 = d\gamma.$$

Hence $\gamma = d^{-1}$ (and $|\langle x_i, y_j\rangle| = d^{-1/2}$).

Our goal is to find mutually unbiased sets of bases with maximal size. How large can a mutually unbiased set of bases be? If P and Q are projections onto lines spanned by two vectors from a set of mutually unbiased bases, then $\langle P, Q\rangle$ is 0, 1 or d^{-1}. The Gram matrix of the projections onto lines spanned by vectors from a set of mutually unbiased bases is

$$G = I_m \otimes I_d + \frac{1}{d}(J_m - I) \otimes J_d.$$

We determine the rank of G by counting its nonzero eigenvalues. The eigenvalues of $(J_m - I) \otimes J_d$ are

eigenvalue	multiplicity
$(m-1)d$	1
0	$m(d-1)$
$-d$	$m-1$

Thus the eigenvalues of $I + \frac{1}{d}(J_m - I) \otimes J_d$ are

eigenvalue	multiplicity
m	1
1	$m(d-1)$
0	$m-1$

Thus $rk(G) = 1 + md - m$ and therefore

$$1 + md - m \leq d^2,$$

from which it follows that $m \leq d + 1$.

Note. If we work in $\mathbb{R}^d$ we get

$$1 + md - m \leq \frac{d^2 + d}{2}$$

and then we find that $m \leq 1 + \frac{d}{2}$.

The columns of a unitary matrix form an orthonormal basis. In fact a matrix H is unitary if and only if its columns form an orthonormal basis. Suppose H and K are unitary then the columns of H and K are unbiased if and only if all entries of H^*K have absolute value $\frac{1}{\sqrt{d}}$. So H^*K is flat and since it is a product of unitary matrices it is unitary. Note that H and K are unbiased if and only if I and H^*K are. Thus each flat unitary matrix gives a pair of unbiased bases in $\mathbb{C}^d$ (matrix, identity).

Suppose the columns of matrices $H_1, \ldots, H_m$ and $K_1, \ldots, K_m$ form mutually unbiased bases in $\mathbb{C}^d$ and $\mathbb{C}^e$ respectively. Then the Kronecker products

$$H_i \otimes K_i$$

give a set of m mutually unbiased bases in $\mathbb{C}^{de}$. (This is very easily verified.) It follows that in any dimension there is a set of at least three mutually unbiased bases.

27.4 Real MUB's

We briefly consider the real case. This received almost no attention prior to the physicists' work on the complex case.

We note first that a flat orthogonal matrix is a scalar multiple of a Hadamard matrix. It follows that if we have a real pair of mutually unbiased matrices in $\mathbb{R}^d$ then either $d = 2$ or $4 \mid d$.

27.4.1 Lemma. *If there is a set of three mutually unbiased bases in $\mathbb{R}^d$, then d is an even square.*

Proof. Suppose H and K are $d \times d$ Hadamard matrices such that the columns of

$$I,\ \frac{1}{\sqrt{d}}H,\ \frac{1}{\sqrt{d}}K$$

are mutually unbiased. Then

$$\frac{1}{d}H^T K$$

must be a flat real orthogonal matrix and therefore

$$\frac{1}{\sqrt{d}}H^T K$$

is a Hadamard matrix. This implies that $\sqrt{d}$ must be rational. □

27.4.2 Lemma. *If there is a set of four mutually unbiased bases in $\mathbb{R}^d$, then $16 \mid d$.*

Proof. Suppose we have four mutually unbiased bases in $\mathbb{R}^d$, the first of which is the standard basis, and assume that $d = 4s^2$. Then the last three bases come from three Hadamard matrices H, K and L such that if x, y and z respectively are columns from these three matrices, then

$$\langle x, y\rangle = \langle x, z\rangle = \langle y, z\rangle = 2s.$$

We consider the equation

$$\langle \mathbf{1}, (x+y) \circ (x+z)\rangle = \langle x+y, x+z\rangle.$$

Since x, y and z are ± 1 vectors, the entries of $x+y$ and $x+z$ are 0 and ± 2 and therefore the left side above is divisible by 4. On the other hand

$$\langle x+y, x+z\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle x, z\rangle + \langle y, z\rangle = 4s^2 \pm 2s \pm 2s \pm 2s$$

and therefore s must be even. □

27.5 Cayley Graphs and Incidence Structures

Let G be an abelian group and suppose $D \subseteq G$. The Cayley graph $X(G, D)$ has the elements of G as its vertices, and (g, h) is an arc if $hg^{-1} \in D$. Any character ψ of G is an eigenvector for $A(X)$, with eigenvalue $\psi(D)$.

The restriction $\psi \restriction D$ lies in $\mathbb{C}^d$, and if ψ and φ are characters of G, then

$$\langle \psi \restriction D, \varphi \restriction D \rangle = (\psi\varphi^{-1})D.$$

Since the product $\psi\varphi^{-1}$ is a character of G, we see that the above inner product is an eigenvalue of $X(G, D)$. In particular the squared cosine of the angle between the complex lines spanned by $\psi \restriction D$ and $\varphi \restriction D$ is the absolute value of an eigenvalue of $X(G, D)$.

If $D \subseteq G$ then we can view the translates Dg, where $g \in G$, as the blocks of an incidence structure. If N is the adjacency matrix of $X(G, D)$, then the adjacency matrix of the incidence graph of this incidence structure can be taken to be

$$A = \begin{pmatrix} 0 & N \\ N^T & 0 \end{pmatrix}.$$

Then

$$A^2 = \begin{pmatrix} NN^T & 0 \\ 0 & N^TN \end{pmatrix}.$$

Now since G is abelian, $NN^T = N^TN$ and hence if θ is an eigenvalue of A, then θ^2 is an eigenvalue of NN^T.

On the other hand, since N is normal, there is an orthogonal basis of $\mathbb{C}^v$ that consists of common eigenvectors of N and N^T. If

$$Nz = \lambda z$$

then

$$N^Tz = \bar{\lambda}z$$

and z is an eigenvector for NN^T with eigenvalue $\lambda\bar{\lambda}$. It follows that

$$\theta = \pm|\lambda|.$$

Since an incidence graph is bipartite, its spectrum is symmetric about zero, and we conclude that the number of negative eigenvalues of the incidence graph is equal to the number of different values taken by the squared cosine of the angles between the v lines in $\mathbb{C}^d$ corresponding to the characters of G.

27.6 Difference Sets

We apply the machinery developed in the previous section to construct sets of $d^2 - d + 1$ equiangular lines in $\mathbb{C}^d$. While this is interesting in its own right, it also serves as a warm up for the more difficult task of constructing mub's.

The *group algebra* of the group G over $\mathbb{C}$ consists of all sums

$$\sum_g c_g g$$

where only finitely many of the coefficients c_g are not zero. (As our groups will be finite, this restriction will not be an issue.) We add and multiply these sums in the obvious fashion. If $D \subseteq G$, we identify D with the element

$$\sum_{d \in D} d$$

of the group algebra. We also use D^{-1} to denote

$$\sum_{d \in D} d^{-1}.$$

Then DD^{-1} can be viewed as the multiset of differences gh^{-1}, where g and h run over the elements of D. (One advantage of the group algebra setup is that we can avoid reference to multisets.)

A subset D of G is a *difference set* if there is an integer λ such that

$$DD^{-1} = |D|1_G + \lambda(G - 1_G).$$

The difference set is parameterized by the triple

$$(|G|, |D|, \lambda);$$

we call λ the *index* of the difference set. This definition of difference set is consistent with the one we used in 5.1, except that there we used abelian groups with addition (rather than multiplication) as the group operation.

If $D \subseteq G$, then we can form an incidence structure with the elements of G as its points and the translates Dg (for g in G) as its blocks. In the cases of interest to us there will be $|G|$ distinct translates of D.

27.6.1 Lemma. *If D is a difference set in the group G with index λ, then the associated incidence structure is a symmetric design with parameter set $(|G|, |D|, \lambda)$.*

Proof. Exercise. □

The incidence matrix of the incidence structure is the adjacency matrix of the Cayley graph $X(G, D)$.

27.7 Difference Sets and Equiangular Lines

Suppose D is a difference set with index λ in the abelian group G and assume $d = |D|$ and $v = |G|$. Then the vectors

$$\psi \restriction D,$$

where ψ runs over the characters of G, span a set of v lines in $\mathbb{C}^d$. To determine the angles, we need the values $|\psi(D)|$. We have

$$|\psi(D)|^2 = \psi(D)\overline{\psi(D)} = \psi(DD^{-1})$$

and since

$$DD^{-1} = d1_G + \lambda(G - 1_G)$$

it follows that

$$|\psi(D)|^2 = d - \lambda + \lambda\psi(G).$$

If ψ is the trivial character, then $\psi(D) = d$ and $\psi(G) = v$ and consequently

$$v = 1 + \frac{d^2 - d}{\lambda}.$$

If ψ is not trivial, then $\psi(G) = 0$ and

$$|\psi(D)|^2 = d - \lambda.$$

This implies that the restrictions $\psi \restriction D$ form a set of equiangular lines in $\mathbb{C}^d$. In particular, when $\lambda = 1$ we obtain a set of $d^2 - d + 1$ equiangular lines in $\mathbb{C}^d$. It can be shown that this set of lines meets the relative bound.

Since the values of a character of an abelian group are roots of unity, the set of lines we obtain from a difference set are spanned by flat vectors. And since he characters form a group, these vectors form a group under Schur multiplication. (This group is a homomorphic image of G. Is it isomorphic to G?)

27.8 Relative Difference Sets and MUB's

Let G be a group with a normal subgroup N. A subset D of G is a *relative difference set* if

$$DD^{-1} = |D|1_G + \lambda(G - N). \tag{27.8.1}$$

Thus a difference set relative to the identity subgroup is a difference set as before.

We offer a relevant example. Let $\mathbb{F}$ be a field of odd order q, let G be the vector space of dimension two over $\mathbb{F}$ and let N be the subgroup $(0, \mathbb{F})$ of G. Then

$$D = \{(x, x^2) : x \in \mathbb{F}\}$$

is a difference set relative to N with index 1.

The defining equation for a relative difference set implies that no two elements of D lie in the same coset of N, and thus wwe have the bound

$$|D| \leq |G : N|.$$

A relative difference set is *semiregular* if equality holds in this bound. Only semiregular relative difference sets will be of interest to us.

If we set d equal to $|D|$ and apply the trivial character to each side of (27.8.1), we find that

$$d^2 = d + \lambda|N|(|G : N| - 1)$$

and consequently if D is semiregular, then

$$d = \lambda|N|, \qquad |G| = \lambda|N|^2.$$

We can divide the characters of G into three classes:

(a) The trivial character ψ, for which $\psi(G) = |G|$.

(b) Non-trivial characters ψ such that $\psi \restriction N$ is trivial, where $\psi(G) = 0$ and $\psi(N) = 0$.

(c) Non-trivial characters ψ such that $\psi \restriction N$ is not trivial, where $\psi(G)$ and $\psi(N)$ are both zero.

We note that the characters whose restriction to N is trivial form a subgroup N_* of the character group G^*, isomorphic to $(G/N)^*$. The corresponding values of $|\psi(D)|^2$ are

(a) $|D|^2$,

(b) $|D| - \lambda|N| = 0$,

(c) $|D|$.

If φ and ψ are characters of G, then $\varphi \restriction D$ and $\psi \restriction D$ are orthogonal if and only if $\varphi\psi^{-1} \restriction N$ is trivial. Hence the characters in a given coset of N_* form an orthogonal basis of $\mathbb{C}^d$, and so we obtain as set of d/λ mutually unbiased bases. We may adjoin the standard basis to this set, thus arriving at a set of $1 + \lambda^{-1}d$ mub's in $\mathbb{C}^d$.

27.9 Type-II Matrices over Abelian Groups

Let N be a group. If W is a matrix with entries from N, we define $W^{(-)}$ to be the matrix with the same order such that

$$(W^{(-)})_{i,j} = W_{i,j}^{-1}.$$

A *type-II matrix over* N is a $v \times v$ matrix W with entries from N such that

$$WW^{(-)T} = vI + \lambda N(J - I).$$

If we apply the trivial character of N to both sides, we obtain the equation

$$J^2 = vI + \lambda|N|(J - I),$$

whence we have

$$v = \lambda N.$$

Suppose N is abelian. If ψ is a character of N, let W_ψ denote the matrix we get by applying ψ to the entries of W. Then $\psi(N) = 0$ if ψ is not trivial and

$$W_\psi W_\psi^* = vI.$$

Thus W_ψ is a flat type-II matrix. If φ is also a character of N, then

$$W_\psi \circ W_\varphi = W_{\psi\varphi}$$

and so the matrices W_ψ form a group of order $|N|$ under Schur multiplication.

27.10 Difference Sets and Equiangular Lines

27.10.1 Lemma. *Let G be an abelian group of order n, and let $\psi_1, \ldots, \psi_n$ be the characters of G. Suppose N is a 01-group matrix over G. If h^T is a row of N, then the number of angles between the lines spanned by the vectors $h \circ \psi_r$ is one less than the number of eigenvalues of NN^T.*

Proof. If ψ and φ are characters of G, then

$$\langle h \circ \psi, h \circ \varphi \rangle = \langle h, \overline{\psi} \circ \varphi \rangle \tag{27.10.1}$$

If χ is a character for G, then $N\chi = \lambda\chi$ for some λ and therefore, since the entries of χ are complex numbers of norm 1,

$$|\langle h, \chi \rangle| = |\lambda|.$$

So

$$|\langle h \circ \psi, h \circ \varphi \rangle|^2$$

is equal to the eigenvalue of NN^T on $\overline{\psi}\varphi$. □

Note that if the weight of the vector h above is d, then the vectors $h \circ \psi$ lie a d-dimensional subspace of $\mathbb{C}^n$.

If X is the incidence graph the design, then

$$A(X) = \begin{pmatrix} 0 & N \\ N^T & 0 \end{pmatrix}$$

and

$$A(X)^2 = \begin{pmatrix} NN^T & 0 \\ 0 & N^TN \end{pmatrix}.$$

It follows that number of angles is equal to the number of non-negative eigenvalues of X.

If N is a group matrix over $\mathbb{Z}_{n^2+n+1}$ and an incidence matrix for a projective plane of order n, then

$$NN^T = nI + J$$

which has eigenvalues $(n+1)^2$ and n (with multiplicity 1 and n^2+n respectively). Hence we obtain a set of n^2+n+1 equiangular lines in $\mathbb{C}^{n+1}$

whenever n is a prime power. The size of this set of lines meets the relative bound, as you are invited to prove.

A complex matrix is *flat* if all its entries have the same absolute value. The vectors spanning the $d^2 - d + 1$ lines are flat; it can be shown that a set of flat equiangular lines in $\mathbb{C}^d$ has size at most $d^2 - d + 1$.

27.11 Affine Planes

Let V be a vector space of dimension two over $GF(q)$; we write it elements as pairs (x, y). Let $[a, b]$ denote the set of points

$$\{(x, y) : y = ax + b\}.$$

This a line in the affine plane over $GF(q)$, and as we vary a and b we get all lines except those parallel to the y-axis—the lines with infinite slope. It is easy to verify that this structure is a divisible semisymmetric design. Our problem is to show that there is an abelian group of automorphisms acting regularly on points and lines.

There are two obvious sets of automorphisms. Let $T_{u,v} : V \to V$ be given by

$$T_{u,v}(x, y) = (x + u, y + v).$$

We call the maps $T_{u,v}$ *translations*, they form an abelian group of order q^2. If (x, y) is on the line $[a.b]$, then

$$y + v - (a(x + u) + b) = (y - ax - b) - (au - v)$$

and therefore $T_{u,v}(x, y)$ is on $[a, b - au + v]$. Thus we can define the image of $[a, b]$ under $T_{u,v}$ to be $[a, b - au + v]$, and with this definition $T_{u,v}$ is an automorphism of our incidence structure. We see that translations are automorphisms that each line to a parallel line. In particular you may show that the group of translations has q orbits on lines.

We can also define *dual translations* $S_{u,v}$ by

$$S_{u,v}[a, b] = [a + u, b + v].$$

Then

$$y - (a + u)x - (b + v) = y + ux + v - ax - b$$

and so $S_{u,v}$ maps lines on (x, y) to the lines on $(x, y + ux + v)$. Again we get a group of automorphisms, with q orbits on points.

What we need though is an abelian group with one orbit on points and one orbit on lines. Define

$$R_{u,v} = T_{u,v}S_{u,0}.$$

Then these q^2 automorphisms from an abelian group of order q^2 that acts transitively on point and on lines. Consequently we get a set of q mutually unbiased bases in $\mathbb{C}^q$, that are all unbiased relative to the standard basis.

This construction does not make use of the fact that finite fields are associative, and we may use a *commutative semifield* in place of a field. All known examples of mub's can be constructed in this way.

27.12 Products

If $H_1, \ldots, H_m$ is a set of unitary matrices describing a set of mub's in $\mathbb{C}^d$ and $K_1, \ldots, K_m$ is a second set giving mub's in $\mathbb{C}^e$, then the products

$$H_i \otimes K_i, \qquad (i = 1, \ldots, m)$$

give us a set of mub's in $\mathbb{C}^{de}$. This may not seem to be a very efficient construction, but in many cases it is the best we can do. If d is a prime power then there is a mutually unbiased set of bases of size $d+1$ in $\mathbb{C}^d$; hence this product construction guarantees the existence of a mutually unbiased set of three bases in any dimension. When $d \cong 2$ modulo four, there is no better bound known in general.

There is one construction, due to Beth and Wocjan, which is better than the product construction in some cases. Suppose we have an $OA(k, q)$ and a flat unitary matrix H of order $q \times q$. Our array can be viewed as an incidence structure with q^2 lines and kq points. Let M be the incidence matrix of the dual; this has order $q^2 \times kq$. Then the kd^2 vectors

$$(\mathbf{1} \otimes He_i) \circ Me_j$$

form k mutually unbiased bases in $\mathbb{C}^{q^2}$. If $q = 26$, then the product construction provides five mub's in $\mathbb{C}^{576}$. There is an $OA(26, 6)$, and so we obtain six mub's in $\mathbb{C}^{576}$.

27.13 Amply-Regular Structures

An incidence structure is *amply regular* if there is a constant μ such that any two vertices at distance two in its incidence graph have exactly μ common neighbors. An amply regular incidence structure where $\mu = 1$ is just a partial linear space.

27.13.1 Lemma. *If $\mathcal{S}$ is a connected amply-regular incidence structure and $\mu > 1$, then it is point and block regular and the number of points equals the number of blocks.*

Proof. See the exercises. □

We say incidence structure is *regular* if its incidence graph is regular. A connected regular amply-regular incidence structure is known as a *semisymmetric design*.

An incidence structure is *point divisible* if we can partition its points into classes such that two points are in the same class if and only if they are not collinear. Dually we have block-divisible structures. If $\mathcal{S}$ is point divisible, any block contains at most one point in each class.

If $\mathcal{S}$ is a semisymmetric design with parameters (v, k, λ), then each point is collinear with exactly $k(k-1)/\lambda$ other points. (You get to prove this in the exercises, but there is a hint.) It follows that in a divisible semisymmetric design, all point classes have the same size. Each block contains at most one point from each class.

27.13.2 Lemma. *A semisymmetric design is point divisible if and only if it is block divisible.*

Proof. Let N be the incidence matrix a point-divisible semisymmetric design with parameters (v, k, λ). Assume that the point classes have size f, and that there are exactly m of them. We have

$$NN^T = kI + \lambda A, \quad N^T N = kI + \lambda B$$

where A and B are the adjacency matrices of graphs X and Y respectively. Since the design is point divisible, X is a complete multipartite graph $\overline{mK_f}$. As NN^T and $N^T N$ are cospectral, Y is cospectral to X and therefore it too must be isomorphic to $\overline{mK_f}$. □

We will refer to the classes of a divisible semisymmetric design as its *fibres*. If $\mathcal{S}$ is divisible with fibres of size f, then

$$NN^T = kI + \lambda(J_m - I_m) \otimes J_f = (k-\lambda)I + \lambda J - \lambda I_m \otimes J_f.$$

We consider a relevant class of examples. Let $\mathcal{A}$ be an affine plane of order q with a parallel class deleted, for example in the classical case delete the lines parallel to the y-axis. Now let $\mathcal{S}$ be the incidence structure formed from the q^2 points of $\mathcal{A}$ and the q^2 lines remaining. This is a divisible regular and amply-regular incidence structure with parameters $(q^2, q, 1)$ and fibres of size q.

27.14 Quotients of Divisible Designs

A divisible semisymmetric design has a natural quotient structure on its fibres, which we investigate here.

27.14.1 Lemma. *Let α and β be disjoint blocks in a point-divisible semisymmetric design. If x is a point on α and C is the point class of x, then β contains exactly one point from C and α is the only block on x disjoint from β.*

Proof. Let s be the number of blocks on x disjoint from β. Since $\alpha \cap \beta = \emptyset$ we have that $s \geq 1$. Count flags (z, γ) where $z \in \beta$ and γ is on x and z. This yields

$$|\beta \setminus C|\lambda = (k-s)\lambda$$

and so the number of points on β collinear with x equals the number of blocks on x that meet β. Since β cannot contain two points in C, we see that $s = 1$. □

This lemma can be used to provide another proof that point-divisible semisymmetric designs are block-divisible. It implies that two blocks in the same block class meet the same k point classes.

27.14.2 Lemma. *Let $\mathcal{S}$ be a divisible semisymmetric (v, k, α)-design, with classes of size f. Two distinct intersecting blocks are incident with the same point classes if and only if each block meets each point class.*

Proof. The number of points collinear with a given point is $k(k-1)/\lambda$ and therefore

$$f = v - \frac{k(k-1)}{\lambda}.$$

If we divide each side of this by f and rearrange, we find that

$$\frac{v}{f} = \frac{k(k-1)}{f\lambda} + 1$$

and so

$$\frac{v}{f} - k = \frac{k(k-1)}{f\lambda} - (k-1) = (k-1)\frac{k - f\lambda}{\lambda}$$

So $v = kf$ if and only if $k = \lambda f$.

If $v = kf$ then each block meets each meets each point class.

Now assume conversely that there are distinct intersecting blocks α and β that meet the same point classes. If θ is a point or a block, let $[\theta]$ denote its class. By the previous lemma, each block in $[\beta]$ meets the same set of point classes, and therefore each point in one of this classes lies in a block from $[\beta]$. Hence each of the k points on α must lie in a block from $[\beta]$ and, since $\alpha \notin [\beta]$, each of these blocks meets α in λ points. Therefore $k = f\lambda$. □

A divisible semisymmetric design where each block meets each point fibre is a *transversal design*.

27.14.3 Corollary. *Let $\mathcal{S}$ be a divisible semisymmetric (v, k, λ) design with fibres of size f. If $k > f\lambda$, then the quotient of $\mathcal{S}$ is a symmetric design with parameters $(v/f, k, \lambda f)$. If $k = \lambda f$, then $\mathcal{S}$ is a transversal design.*

Suppose $\mathcal{S}$ is a divisible semisymmetric design with $v = kf$. If we set

$$F = \{1, \ldots, f\}$$

then we can represent each block by a k-tuples of elements from F. This set of k-tuples forms an orthogonal array with index λ, that is, an $OA_\lambda(f, k)$. If $\lambda = 1$, we have an $OA_1(k, k)$. An affine plane gives an $OA_1(k, k+1)$, and it is not hard to show that any $OA_1(k, k)$ can be extended to an affine plane.

27.15 Incidence Graphs

If $\mathcal{S}$ is a divisible semisymmetric design with parameters (v, k, λ) and with fibres of size f, then

$$NN^T = kI + \lambda(J_m - I_m) \otimes J_f = (k - \lambda)I + \lambda J - \lambda I_m \otimes J_f.$$

We compute the eigenvalues of this matrix. There are three relevant subspace of $\mathbb{R}^v$. First, the constant vectors. These form a 1-dimensional eigenspace with eigenvalue k^2. Next we have the subspace formed by the vectors that are constant on fibers and sum to zero. If z is such a vector, then $(I_m \otimes J_f)z = fz$

$$NN^T z = (k - \lambda f)z$$

and we have an eigenspace with dimension $\frac{v}{f} - 1$. Finally the vectors which sum to zero on each fibre form an eigenspace of dimension $v - \frac{v}{f}$ with eigenvalue $k - \lambda$.

27.15.1 Theorem. *Let $\mathcal{S}$ be a divisible semisymmetric design with parameters (v, k, λ) and fibres of size f. If $v = kf$ and $\mathcal{S}$ has an abelian group G of automorphisms, then the standard basis of $\mathbb{C}^k$ together with the vectors formed by the restriction of a character of G to a block form a set of $1 + \frac{v}{f}$ mutually unbiased bases in $\mathbb{C}^k$.* □

We find examples of structures that satisfy these conditions.

Part IV

Tools

Chapter 28

Permutation Groups

We offer an introduction to some of the theory of permutation groups.

28.1 Permutation Groups and Representations

A *permutation group* on the set X is a subgroup of $\mathrm{Sym}(X)$. A *permutation representation* of the group G is a homomorphism from G into $\mathrm{Sym}(X)$. The image in $\mathrm{Sym}(X)$ of G is isomorphic to a quotient group of G; if it is isomorphic to G itself then we say that the representation is *faithful*. If g is a permutation of X and $Y \subseteq X$ then we define the set Yg by

$$Yg := \{ig : i \in Y\}.$$

We call Yg a *translate* of Y under the action of G. If $Yg = Y$ we say that Y is *fixed* by g. If $Yg = Y$ for all elements g in the permutation group G then we say similarly that Y is fixed by G. The set Y is fixed by g if and only if it is fixed by the subgroup $\langle g \rangle$. Observe that, having defined Yg for each subset Y of G and each element g of G, we have described a permutation representation of G on the power set of X.

If $A \subseteq G$ and $i \in X$ then

$$iA = \{ia : a \in A\}.$$

A subset S of X is an *orbit* of G if it is fixed by G and, given any two elements i and j of S, there is an element g in G such that $ig = j$. We

could equivalently define an orbit to be a minimal fixed subset of X. We will also refer to the orbits of an element g, when strictly we mean the orbits of $\langle g \rangle$. Any fixed subset of X is a disjoint union of orbits. An orbit can consist of just one element, in which case it is called a *fixed point*. A group with just one orbit is said to be *transitive*. If G is permutation group on X then its elements are functions defined on X. Hence if $g \in G$ and $Y \subseteq X$ then the restriction $g \restriction Y$ is defined. If Y is an orbit of G then $g \restriction Y$ is a permutation of Y, and the set

$$\{g \restriction Y : g \in G\}$$

is a permutation group on Y. We denote this group by $G \restriction Y$, and observe that the mapping $g \mapsto g \restriction Y$ is a homomorphism from G into $\mathrm{Sym}(Y)$. Thus it is a permutation representation of G, and generally not a faithful one.

If G and H are permutation groups acting respectively on the disjoint sets X and Y then $G \times H$ acts on $X \cup Y$ according to the rule

$$(g, h) : i \mapsto \begin{cases} ig, & \text{if } i \in X; \\ ih, & \text{if } i \in Y. \end{cases}$$

This gives us a permutation group on $X \cup Y$ which we could call the *sum* of G and H. Both X and Y are fixed sets for this group. We can also define an action of $G \times H$ on $X \times Y$ by insisting that

$$(g, h) : (i, j) \mapsto (ig, jh).$$

This gives a permutation group on $X \times Y$, which is called the *product* of X and Y. The product of G and H is transitive if and only if G and H are both transitive; the sum of G and H is never transitive. It is easy to extend these definitions to sum and products of any number of permutation groups.

There are a number of ways of constructing permutation groups from a given abstract group. In a particular a given group G acts on its underlying set of elements by *conjugation*. This associates to an element g in G the mapping such that, if $x \in G$ then

$$x \mapsto x^g = g^{-1}xg.$$

It is easy to check that this is a permutation representation of G, which is faithful if and only if the *centre* $Z(G)$ is trivial, and is never transitive. An

orbit of G in this case is known as a *conjugacy class*. More generally we can regard G as acting on the set of all subsets of G: if $A \subseteq G$ and $g \in G$ then define A^g to be

$$\{a^g : a \in A\}.$$

If A and B are subsets of G and $B = A^g$ for some element g of G then we say A and B are *conjugate* subgroups of G. A group also acts on its subsets by *left multiplication*: assign to each element g of the mapping λ_g such that

$$\lambda_g : S \mapsto gS = \{gx : x \in S\}.$$

We also have the representation of G on its subsets by right multiplication; here ρ_g is given by

$$\rho_g : x \mapsto xg^{-1}.$$

The action of G on its one-element subsets by left and multiplication are known respectively as the *left* and *right regular representations* of G. They are both transitive and faithful. If S is a subgroup of G then its orbit under right multiplication by elements of G consists of the right cosets of S. The representation arising by restriction to this orbit will be faithful if and only if S contains no normal subgroup of G.

28.2 Counting

Let G be a permutation group on X. If $i \in X$ then the *point-stabilizer* G_i is the subgroup of G formed by the elements which fix i. If $i_1, \ldots, i_r$ are distinct elements of X then

$$G_{i_1,\ldots,i_r} := \bigcap_{j=1}^{r} G_{i_j},$$

i.e., it is the subgroup of elements which fix each point in S. If $g \in G$ then

$$\operatorname{fix} g = \{i \in X : ig = i\}$$

and, for any subset A of G,

$$\operatorname{fix} A = \bigcap_{a \in A} \operatorname{fix} a.$$

Of course, $\operatorname{fix} g = \operatorname{fix} \langle g \rangle$ and, more importantly, $\operatorname{fix} A^g = \operatorname{fix} Ag$.

28.2.1 Lemma. *Let G be a permutation group on the set X and let Ω be an orbit containing the point i. Then there is bijection between the right cosets of G_i in G and the elements of Ω. Furthermore, $j \in \Omega$ if and only if G_j is conjugate to G_i in G.*

Proof. Suppose that $g \in G$ and $ig = j$. Then $ih = j$ if and only if $igh^{-1} = i$, i.e., if and only if $hg^{-1} \in G_i$. But $hg^{-1} \in G_i$ if and only if $h \in G_i g$, which proves our first claim. The second claim follows from the observation that $g^{-1}G_i g = G_{ig}$. □

The first part of this lemma is often called the "orbit-stabilizer relation". It can be expressed in the form

$$|iG| = |G : G_i|.$$

Equivalently, the length of an orbit of G is equal to the index of the stabilizer of an element from the orbit. This lemma also shows that every transitive permutation group can be obtained by considering its action on the cosets of some subgroup by right multiplication. If the group G acts on its set of elements by conjugation and $x \in G$ then the stabilizer of x is known as the *centralizer* of x. It is denoted by $C_G(x)$. When G acts on the conjugates of a subgroup H then the stabilizer of H is called the *normalizer* of H, and denoted by $N_G(H)$. From the second part of ??Lemma 3.1, we see that the latter representation can be obtained by considering the action of G on the cosets of $N_G(H)$ by right multiplication.

28.2.2 Lemma. *The number of orbits of a permutation group is equal to the average number of points fixed by an element of the group.*

Proof. Let G be a permutation group on X. Consider the set $\mathcal{P}$ of ordered pairs (i, g), where $i \in X$ and $g \in G_i$. Then

$$|\mathcal{P}| = \sum_{g \in G} |\operatorname{fix} g|.$$

Assume that X is the disjoint union of orbits $X_1, \ldots, X_r$. Then we also have

$$|\mathcal{P}| = \sum_{j=1}^{r} \sum_{i \in X_j} |G_i|. \tag{28.2.1}$$

The cardinality of G_i is independent of the choice of i in X_j, by the previous lemma. Hence $\sum_{i \in X_j} |G_i|$ equals $|X_j|\,|G_i|$ for some i in X_j. Since $|X_j| =$

$|G : G_i|$, the inner sum in (28.2.1) is equal to $|G : G_i|\,|G_i|$, and so the right side of (28.2.1) is just $|G|$ times the number of orbits of G. It follows that the number of orbits of G is equal to $|G|^{-1} \sum_{g \in G} |\operatorname{fix} g|$, which is what we claimed. □

We now consider some of the consequences of the previous results, starting with ??Lemma 3.1. Consider the case where G is a p-group, acting by conjugation on its set of elements. By ??Lemma 3.1, each orbit of G must have prime power length. Also there is at least one orbit of length one, since the identity element e is conjugate only to itself. It follows that that the number of fixed points is non-zero, and divisible by p. Hence there is an element x of G such that $G = C_G(x)$. In other words the centre $Z(G)$ of a p-group is always a non-trivial subgroup.

As a more complicated example, we establish the existence of Sylow p-subgroups. Let G be a group of order $n = mp^k$, where $(m, p) = 1$. Let $\mathcal{P}$ the be the set of all subsets of G with cardinality p^k. Then G acts on this set by right multiplication. If $\beta \in \mathcal{P}$ then β is a union of left cosets of its stabilizer G_β. Hence $|G_\beta|$ must divide $|\beta| = p^k$ and so, by the orbit-stabilizer relation, $|\beta G|$ is divisible by m. If $|\beta G| = m$ then $|G_\beta|$ is a subgroup of G with order p^k, i.e., it is Sylow p-subgroup. If $|\beta G| > m$ then it must be divisible by p. Thus we see that either G has a Sylow p-subgroup, or else all orbits of G on $\mathcal{P}$ have length divisible by p. But the latter is impossible, because

$$|\mathcal{P}| = \binom{mp^k}{p^k}$$

is congruent to m modulo p. (The proof of this is straightforward.)

Finally we present a simple application of Burnside's lemma.

28.2.3 Lemma. *Let G be a transitive permutation group on the set X. If $|X| > 1$, there is an element of G with no fixed points.*

Proof. Suppose $|X| = n$. The identity element of G fixes n points, and so if all the other elements of G have fixed points then the average number of points fixed by an element of G is greater than one. Hence G must have at least two orbits, and cannot be transitive. □

If $g \in G$ with exactly m fixed points then it fixes exactly m^k points from X^k. By Burnside's lemma then, the number of orbits of G on X^k is equal

to

$$|G|^{-1} \sum_{g \in G} |\operatorname{fix} g|^k.$$

When G is transitive on X, this is equal to the number of orbits of a point-stabilizer G_x on X^{k-1}. (The proof of this is left as an exercise.) If G is transitive on X then the number of orbits of G on X^2 is known as its *rank*.

28.3 Transitivity

Let G be a permutation group on X and let U be a subgroup of G. If $i \in \operatorname{fix} U$ and $g \in N_G(U)$ then

$$igU = igUg^{-1}g = iUg = ig.$$

Hence $\operatorname{fix} U$ is fixed by g. Thus $\operatorname{fix} U$ is a union of orbits of $N_G(U)$. Our next result, due to Jordan, determines when it is a single orbit.

28.3.1 Theorem. *Let G be a transitive permutation group on the set X, let x be an element of X and let U be a subgroup of G_x. Assume that, if $g \in G$ such that $U^g \leq G_x$, there is an element h of G_x with $U^g = U^h$. Then $N_G(U)$ acts transitively on $\operatorname{fix} U$.*

Proof. Let y be a second point fixed by U. (If there is no such y, there is nothing to prove.) Since G is transitive, it contains an element g that $x = yg$. Then $xg^{-1}Ug = yUg = yg = x$, whence U^g fixes x. Thus $U^g \leq G_x$ and so, by our hypothesis, there is an element h of G_x such that $U^h = U^g$. From this we see that $U^{gh^{-1}} = U$ and therefore $gh^{-1} \in N_G(U)$. Now $ygh^{-1} = xh^{-1} = x$, and so we have found that there is an element of $N_G(U)$ sending y to x. It follows immediately that $N_G(U)$ acts transitively on $\operatorname{fix} U$. □

Two important cases where this result applies are when $U = G_x$, and when U is a Sylow p-subgroup of G_x, for some prime p.

A permutation group G acting on X is said to be *regular* if it is transitive, and any element which fixes a point is the identity. Thus all point-stabilizers of a regular group are trivial and so, by the orbit-stabilizer relation, $|G| = |X|$. Conversely, if G is transitive and $|G| = |X|$ then G is regular. The left and right regular representations of group give regular permutation groups. It can be shown that $N_G(G_x)$ acts regularly on $\operatorname{fix} G_x$.

28.3.2 Lemma. *(Gleason). Let G be a permutation group on X and let p be a prime. Suppose Y is a subset of X such that, for each point y in Y there is a p-subgroup of $P(y)$ of G which fixes y, but no other point of X. Then Y is contained in an orbit of G.*

Proof. Assume $y \in Y$ and let $P(y)$ be as above. Under the action of $P(y)$, the G-orbit yG divides into orbits. With the exception of the orbit formed by y itself, each of these has length divisible by p. Thus $|yG| \equiv 1$, modulo p. Now suppose that $z \in Y$, but not in yG. As z is the unique fixed point of $P(z)$, it follows that $P(z)$ fixes no point in yG. Hence each $P(z)$-orbit contained in yG has length divisible by p, and so $|yG|$ must be divisible by p. This contradicts our previous conclusion, and forces us to conclude that $Y \subseteq yG$. □

Gleason's lemma implies that the Sylow p-subgroups of a group are conjugate. To see this, let G be a group and let $\mathcal{P}$ be the set formed by its Sylow p-subgroups. The G acts on $\mathcal{P}$ by conjugation. Any Sylow p-subgroup of G fixes a unique element of $\mathcal{P}$ (namely itself) (why?), its remaining orbits all have length divisible by p. Thus we can apply Gleason's lemma to deduce that G acts transitively on $\mathcal{P}$, and hence that the Sylow p-subgroups are conjugate. (Note also that since $|\mathcal{P}| \equiv 1$ modulo p, any p-subgroup of G fixes at least one element of $\mathcal{P}$. This implies that each p-subgroup lies in a Sylow p-subgroup.) Gleason's lemma can in turn be derived from the fact that the Sylow p-subgroups of a group are conjugate.

We can now show that if G is a transitive group and U is a Sylow p-subgroup of G_x then it satisfies the hypothesis of Jordan's Theorem. For if U is a Sylow p-subgroup of G_x and $U^g \leq G_x$ then U^g is also a Sylow p-subgroup of G_x. Thus it is conjugate to U in G_x.

28.4 Higher Transitivity

If G is a permutation group on the set X the we can define an action of G on the set X^k off ordered k-tuples as follows. If $(x_1, \ldots, x_k) \in X^k$ and $g \in G$ then

$$(x_1, \ldots, x_k)g := (x_1 g, \ldots, x_k g).$$

Thus we obtain a collection of representations of G, one for each value of k. Since no k-tuple of distinct elements can be mapped to a k-tuple with a repeated element, these representations will not be transitive when $k > 1$,

even if G is. A permutation group on X is said to be *k-transitive* if it acts transitively on the ordered k-tuples of distinct elements from X. We see immediately that, if $k > 1$, a k-transitive group is also $(k-1)$-transitive.

28.4.1 Lemma. *A group G acting on X is k-transitive if and only if it is transitive, and the stabilizer of any point x of X is $(k-1)$-transitive on $X \setminus x$.*

Proof. Exercise. □

Following the classification of the finite simple groups, it possible to list all the 2-transitive permutation groups. It is too soon for us to say much about them, but we can give some examples. The symmetric and alternating groups on n letters are, respectively, n- and $(n-2)$-transitive. The general linear group $GL(n, \mathbb{F})$ induces a 2-transitive group of permutations of the points of $PG(n-1\mathbb{F})$. If $X = \mathbb{F} \cup \infty$ then the group formed by the mappings

$$x \mapsto \frac{ax+b}{cx+d},$$

with a, b, c and d from $\mathbb{F}$ and $ad-bc \neq 0$, is 3-transitive. (Here one operates with ∞ the way a calculus student would always like to. Thus $1/\infty = 0$, $\infty/\infty = 1$, etc.) Apart from the symmetric and alternating groups there are only two further 4-transitive groups and two 5-transitive groups. (These are the Mathieu groups.)

28.5 Homogeneity

Instead of considering the action of G on X^k, we may also look at its action on the set $\binom{X}{k}$ of all k-subsets of X. (For combinatorial purposes, this is often more natural.) A permutation group is *k-homogeneous* if it acts transitively on $\binom{X}{k}$. If $|X| = n$ then G is k-homogeneous if and only if it is $(n-k)$-homogeneous. Our first difficulty appears at once. When is it true that a k-homogeneous group, is also $(k-1)$-homogeneous?

28.5.1 Lemma. *(Wielandt). Let G be a k-homogeneous group on X, with $k \geq 2$. Suppose that if q is a prime power dividing k then $q \leq |X| + 1 - k$. Then G is $(k-1)$-homogeneous.*

Proof. Let Ω be an orbit of G in its action on $\binom{X}{k-1}$. Since G is k-homogeneous, each k-set from X must contain the same number of $(k-1)$-sets from Ω. Let this number be ℓ; it will suffice to show that $\ell = k$. Choose an integer s such that $k \leq s \leq |X|$, a subset S of X with size s and consider the ordered pairs (A, B), where $A \in \Omega$, $B \in \binom{S}{k}$ and $A \subseteq B$. There are $\binom{s}{k}\ell$ such pairs. Since every set A lies in exactly $(s-k+1)$ different sets B, it follows that $s-k+1$ divides $\binom{s}{k}\ell\ell$. This must hold for all integers s between k and $|X|$; in particular we may choose s equal to $k+q-1$, for some prime power q dividing k. Then, if $m = 1 + (k/q)$, we find that q divides

$$\binom{k+q-1}{k}\ell = \binom{k+q-1}{q-1}\ell = \binom{qm-1}{q-1}\ell.$$

The last binomial coefficient is not divisible by p (another combinatorial exercise), hence q must divide ℓ. Consequently k divides ℓ and so $k = \ell$ as required. □

If $|X| = n$ and $2k \leq n$ then every prime power dividing k satisifies the condition of the lemma, and so it follows that a group acting k-homogeneously on X must also act $(k-1)$-homogeneously. As a somewhat more bizarre, and less important example, consider a 2-homogeneous group of degree 8. Such a group is 6-homogeneous, by the lemma it is 5-homogeneous, and hence also 3-homogeneous. It was proved by Livingstone and Wagner that if $k \geq 2$ then a k-homogeneous group is $(k-1)$-transitive, and is even k-transitive if $k \geq 5$. (They also proved that k-homogeneous groups are $(k-1)$-homogeneous.)

We prove one result in this direction after the following preliminary.

28.5.2 Lemma. *Let G be a permutation group on X and let P be a Sylow p-subgroup of G. If $x \in X$ and q is power of p dividing $|xG|$ then q divides $|xP|$.*

Proof. We have

$$|xG||G_x : P_x| = |G : G_x||G_x : P_x| = |G : P_x| = |G : P||P : P_x| = |G : P||xP|.$$

Since $|G : P|$ is not divisible by p, we see that q must divide $|xP|$. □

28.5.3 Lemma. *Let G be a k-homogeneous group on a set X of n points, where $2k \leq n$. If there is a non-identity element of G fixing each point in a subset of X of size $k-1$ then G is $(k-1)$-transitive.*

Proof. Let T be a subset of X with size $k-1$. Since G is k-homogeneous, $|G_{T\cup x}|$ is independent of the choice of x in $X\setminus T$. It follows that all orbits of G_T on $X\setminus T$ have same length, ℓ say. Since G_T is not the trivial subgroup, $\ell > 1$. Let p be a prime divisor of ℓ and let P be a Sylow p-subgroup of G_T. By the previous lemma we obtain that all orbits of P on $X\setminus T$ have length divisible by p and hence that $\operatorname{fix} P = T$. By a simple extension of Gleason's lemma, which we have left as exercise, it follows that G is $(k-1)$-transitive. □

28.5.4 Corollary. *Let G act k-homogeneously on the set X. If $|X| \geq k! + k - 1$ then G is $(k-1)$-transitive.*

Proof. Assume $|X| = n$. If G is k-homogeneous then $\binom{X}{k}$ is a single orbit and so $\binom{n}{k}$ must divide $|G|$. Suppose that Y is a subset of $\binom{X}{k-1}$ such that G_Y is trivial. Then the orbit of Y under the action of G (on $(k-1)$-tuples of distinct elements of X) has length equal to $|G|$, and therefore $|G|$ cannot be greater than n^{k-1}. Thus we have

$$\frac{n_{(k)}}{k!} \leq |G| \leq n_{(k-1)},$$

which implies that $n - k + 1 \leq k!$. Given our hypothesis, it follows from this that G_Y cannot be trivial, and so by the previous lemma G must be $(k-1)$-transitive. □

28.6 Primitivity and Imprimitivity

Let G be a transitive permutation group on the set X. A non-empty subset S of X is a *set of imprimitivity* for G if any two translates of S are either equal or disjoint. (Such sets are often called "blocks", but this would be inconvenient for us.) There are two trivial cases; if S is a singleton or the entire set. A group is *imprimitive* if there is a non-trivial set of imprimitivity, and is otherwise *primitive*. The group G acts as a permutation group on the distinct translates of S. The set of translates of a set of imprimitivity will be called a *system of imprimitivity*. Since the translates of S partition X, it follows that $|S|$ divides $|X|$. Thus every transitive permutation group of prime degree is primitive. It is an easy exercise to show that every 2-transitive group is primitive. There are two important characterisations of

primitive groups, the first of which we present now. (The second will be ??Theorem 8.2.)

28.6.1 Lemma. *Let G be a transitive permutation group. Then G is primitive if and only if the stabilizer of a point is a maximal subgroup of G.*

Proof. Suppose that S is a non-trivial set of imprimitivity for G and that $1 \in S$. If $g \in G_1$ then $1 \in Sg$ and therefore $S = Sg$. Thus G_1 is contained in the subgroup H of G formed by the permutations which fix S as a set. If $x \in S$ not equal to 1 then $1h = x$ for some element h of G. Then $x \in Sh \cap S$ and so $Sh = S$. Hence G_1 is a proper subgroup of H. Conversely, if $G_1 < H$ and $H < G$ then $S = 1H$ is a set of imprimitivity for G. For suppose that $g \in G$ and $1h \in 1Hg$ for some element h of H. Then $1hg^{-1} = 1h'$, where $h' \in H$, implying that $h'gh^{-1} \in G_1$. As $G_1 \leq H$, this implies that $g \in H$ and hence that $1Hg = 1H$. Accordingly the translates of $1H$ are equal or disjoint. Since

$$|1H| = |H : G_1| < |G : G_1|,$$

we see that $1H \neq X$ and as $G_1 < H$, we also see that $1H$ is not a singleton. Hence it is a non-trivial set of imprimitivity. □

If A and B are subsets of a group G then

$$AB = \{ab : a \in A, b \in B\}.$$

Even if A and B are subgroups of G, the product set AB is not generally a subgroup. In fact, if $A, B \leq G$ then $AB \leq G$ if and only if $AB = BA$. Hence if $A \trianglelefteq G$ and $B \leq G$ then $AB \leq G$.

28.6.2 Lemma. *Let G be a transitive permutation group on the set X and let 1 be a point in X. A subgroup H of G is transitive if and only if $G_1H = G$.*

Proof. The subgroup H is transitive if and only if for each point i in X there is an element h_i in H such that $1h_i = i$. This is equivalent to requiring that H contain a complete set of coset representatives for G_1 in H, and this yields the lemma. □

A permutation group is said to be $\frac{1}{2}$-*transitive* if its orbits all have the same length. More generally, it is $t\frac{1}{2}$-transitive if it is t-transitive and the stabilizer of t points is $\frac{1}{2}$-transitive.

28.6.3 Lemma. *Let G be a transitive permutation group on X and let N be a normal subgroup of G. If N is not transitive on X then it is $\frac{1}{2}$-transitive, and its orbits form a system of imprimitivity for G.*

Proof. If N is not transitive then $G_1N \neq G$. Since $N \trianglelefteq G$, the product G_1N is thus a subgroup of G, strictly contained in G. If $G_1N = G_1$ then $N \leq G_1$. Then $1N = 1$ and, for any element g of G,

$$1gN = 1gNg^{-1}g = 1Ng = 1g.$$

Thus every point in X is fixed by N and so $N = \langle e \rangle$. If $G_1 < G_1N$ then $1N$ is a set of imprimitivity for G. It is not hard to show that the remaining orbits of N are translates of this, and hence that they all have the same size. □

Thus normal subgroups of transitive permutation groups can provide sets of imprimitivity. A second source is provided by the next result.

28.6.4 Theorem. *(Witt). Let G be a transitive permutation group on the set X and let 1 be a point in X. Let U be a subgroup of G_1 such that if $g \in G$ and $U^g \leq G_1$ then $U^g = U$. Then $\operatorname{fix} U$ is a set of imprimitivity for G.*

Proof. Denote $\operatorname{fix} U$ by F. If $1g \in F$ then $1gU = 1g$ and so $1gUg^{-1} = 1$. Hence $U^{g^{-1}} \leq G_1$ and so $U^{g^{-1}} = U$. Consequently

$$Fg = (\operatorname{fix} U)g^{-1} = \operatorname{fix} U^{g^{-1}} = \operatorname{fix} U = F.$$

This shows that F is a set of imprimitivity for G. □

By way of example, consider the cube. This graph has the property that, for each vertex i in it, there is a unique vertex i' at distance three from it. If G is the automorphism group of the cube, we can thus deduce that G_i fixes i and i', but no other vertices. Thus $\{i, i'\}$ is a set of imprimitivity. A subgroup U satsifying the condition of Witt's theorem is said to be *weakly closed* in G_1. Any weakly closed subgroup U of G_1 satisfies the conditions of Jordan's theorem (??Theorem 4.1), that is, any conjugate of U contained in G_1 is conjugate to U using an element of G_1. A weakly closed subgroup of G_1 is necessarily normal, although not all normal subgroups need be weakly closed.

28.7 Generously Transitive Permutation Groups

A permutation group G on the set X is *generously transitive* if each pair of distinct points of X is swapped by some element of G. The dihedral group of order $2n$ acting in the natural fashion on a set of n points provides a simple example. Generously transitive permutation groups are closely related to association schemes.

28.7.1 Lemma. *Let G be a generously transitive permutation group of rank $d+1$ on the set X. Then the adjacency matrices of the orbitals of G form an association scheme with d classes.*

Proof. Let $\Omega_0, \Omega_1, \ldots, \Omega_d$ be the orbitals of G, with respective adjacency matrices $A_0 = I, A_1, \ldots, A_d$. We only have to verify that A_iA_j is a linear combination of these adjacency matrices, for all i and j. If x and y are points in X then the xy-entry of A_iA_j is equal to the number, p_{ij} say, of points z such that $(x,z) \in \Omega_i$ and $(z,y) \in \Omega_j$. If (x',y') lies in the same orbital as (x,y) then, since there is an element of G mapping (x,y) to (x',y'), p_{ij} is also equal to the number of points z such that $(x',z) \in \Omega_i$ and $(z,y') \in \Omega_j$. This implies that A_iA_j is an integral linear combination of $A_0, \ldots, A_d$. □

The group G in the previous result can be viewed as a group of permutation matrices. If P is one of these matrices then $P^TA_iP = A_i$ and so $A_iP = PA_i$. Thus P commutes with each of the matrices A_i and it can actually be shown that the linear span of these matrices is the vector space of all matrices which commute with each matrix in G. The next result is a simple criterion for generous transitivity.

28.7.2 Lemma. *(Shult). Let G be a transitive permutation group on X. If G contains an involution with exactly one fixed point then it is generously transitive.*

Proof. Let $X' = X \cup \infty$. Let T be the set of all involutions in G with exactly one fixed point. Each element of T determines a 1-factor in the complete graph K' with vertex set X'. Let K be the complete graph with vertex set X. Suppose that s and t are elements of T fixing the points x and y respectively. The graph F' formed by the union of the edges of the 1-factors corresponding to s and t is a disjoint union of even cycles. Hence the

subgraph F of K obtained by deleting the vertex ∞ from F' is the disjoint union of even cycles, and a single path with an odd number of vertices in it. This path has x and y as its endpoints, and its edges come alternately from s and t. The vertices of this path form an orbit for $D = \langle x, y \rangle$ with odd length, m say. The group D is dihedral, with order $2m$. It can viewed as the automorphism group of a cycle with odd length, and a moment's thought shows that it is generously transitive. Thus there is an element of D interchanging x and y, and hence there an element of G which does this. □

The above proof actually shows that if s fixes x and t fixes y then then there is an involution in $\langle s, t \rangle$, conjugate to s and to t, and interchanging x and y. (This also implies that the involutions in G with exactly one fixed point are all conjugate.)

Exercises

(28.2.1)

1. Show that every permutation group is a subgroup of a sum of transitive permutation groups.

2. Give an example of a permutation group with no fixed points and such that every element fixes at least one point.

3. Prove that G acts k-transitively on X if and only if it acts transitively, and the stabilizer of any point x of X acts $(k-1)$-transitively on $X \setminus x$.

4. A permutation group acting on X is *k-closed* if it is not contained in a larger subgroup of $\mathrm{Sym}(X)$ having the same orbits on X^k. Show that a group is 1-closed if and only if it is the sum of some number of symmetric groups.

5. Show that the automorphism group of a directed graph is 2-closed.

6. Show that the collineation group of a projective plane is 3-closed.

7. Let G be a permutation group on the set X. Show that if G is abelian then its 2-closure is abelian, with the same exponent. Show that any prime which divides the order of the 2-closure of G must divide the order of G. (Here G might not be abelian.)

8. Show that if G acts transitively on X and $x \in X$ then the number of orbits of G on X^k is equal to the number of orbits of G_x on X^{k-1}.

9. Let G act transitively on X, let x be an element of X and and let U be a subgroup of G_x. Suppose that the conjugates of U contained in G_x fall into exactly t conjugacy classes under the action of the elements of G_x. Show that $\operatorname{fix} U$ falls into exactly t orbits under the action of $N_G(U)$.

10. Prove the following extension of Jordan's theorem. Let G be a t-transitive permutation group on X and let U be a subgroup of G_1 $(1 \in X)$ such that if $U^g \leq G_1$ for some element g of G then $U^g = U^h$ for some element h of G_1. Then $N_G(U)$ acts t-transitively on $\operatorname{fix} U$.

11. Show that a transitive permutation group is regular if and only if the stabilizer of a point is normal. Hence show that if G is a transitive permutation group, then $N_G(G_x)$ acts regularly on $\operatorname{fix} G_x$.

12. Let G be a permutation group on the set X. If, for each t-subset T of G, there is a p-subgroup P of G with $\operatorname{fix} P = T$, show that G is t-transitive. (This is an extension of Gleason's lemma.)

13. If A and B are subgroups of G, show that $|AB| = |A||B|/|A \cap B|$. Show also that AB is a subgroup of G if and only if $AB = BA$.

14. Let G be a transitive permutation group on X and S be an orbit of subgroup of G which contains the point 1 of X. Show that the intersection of the translates of S which contain 1 is set of imprimitivity for G. (What happens if S is an arbitrary proper subset of X?)

15. Let H and K be subgroups of G. A *double coset* is a subset of the form HgK. Show that the distinct double cosets with respect to H and K formed as g ranges over the elements of G partition G. If G acts transitively on X and $x \in X$, show that the distinct double cosets $G_1 g G_1$ correspond to the orbits of G_1.

16. Let G be transitive group on the set X and let N be a normal subgroup of G. Show that the number of orbits of N on X divides $|G : N|$.

17. Let G be a p-group. If $H < G$, show that $H < N_G(H)$.

18. Show that the number of Sylow p-subgroups of a group is congruent to 1 modulo p.

19. Let H and K be subgroups of G such that $G = HK$. Show that there is a Sylow p-subgroup P of H and a Sylow p-subgroup Q of K such that PQ is a Sylow p-subgroup of G. (This is proved in the text, under the assumption that $K \trianglelefteq G$.)

20. Let G be a transitive permutation group such that G_1 is cyclic. Show that any subgroup of G_1 is weakly closed, and determine when it is strongly closed. (Hint: if C is cyclic and $K \leq C$ show that K consists of all elements in C with order dividing $|K|$.)

21. Let X and Y be orbits of the group G and let d be the greatest common divisor of their lengths. If $x \in X$, show that all orbits of G_x contained in Y have length divisible by d.

22. Let G be transitive group on the set X. Show that, if no two orbits of G_1 have same length, G is generously transitive. (Is this still true if we only require that no two non-trivial orbits have the same length?)

23. Let G act generously transitively on the set X, and let S be a set of imprimitivity for G. Show that $G_{\{S\}}\restriction S$ is generously transitive, and the permutation group induced by G on the translates of S is also generously transitive.

Notes

Biggs and White [BigWh] provide a group-theoretic construction of the Mathieu groups, and an introduction to the Higman-Sims group. (This is one of the sporadic simple groups.) The book by Burnside [Burn] is, naturally enough, written in an outdated style. However it is well worth careful study. Huppert and Blackburn [HuppBl] provide a considerable amount of information, but it is less immediately accessible in that it makes use of character theory. (The first volume of this book, written by Huppert alone, also contains considerable information on permutation groups, in German.) Passman's book [Pass] is elegantly written, and is devoted to classifying the sharply t-transitive permutation groups (with $t \geq 2$). Tsuzuku [Tsuz] presents a reasonable amount of information about groups acting on designs and geometries, but his presentation often follows the original papers very closely. Wielandt's book [Wiel] is the standard reference. Most papers

on permutation groups contain claims of the type “We use the notation of Wielandt [n]”. (Fortunately these claims are generally false.)

Chapter 29

2-Transitive Groups

We introduce Frobenius groups, and obtain some information on the sharply 4-transitive groups.

29.1 Frobenius Groups

We call a group *sharply t-transitive* if it is t-transitive and the stabilizer of any set of t points is the identity. (Thus a sharply 1-transitive group is the same as a regular group.) A group is *elementary abelian* if it is abelian and all non-identity elements in it have the same order. Such a group is necessarily isomorphic to $\mathbb{Z}_p^n$ for some prime p and integer n; equivalently it can be regarded as a vector space over $GF(p)$.

29.1.1 Lemma. *Let G be a sharply 2-transitive group on the set X. Then G has a regular normal subgroup which is an elementary abelian p-group, and $|X|$ is a power of p.*

Proof. Set n equal to $|X|$. Since G is sharply 2-transitive, it has order $n(n-1)$, and any point-stabilizer has order $n-1$. Let γ be an element of G with order p, for some prime p dividing n. As p cannot divide $n-1$, we see that γ has no fixed points. Any two distinct point-stabilizers have only the identity in common, since it is the only element fixing two points. It follows that there are $n(n-2)$ elements in G with exactly one fixed point, and hence $n-1$ elements with none. Suppose $\alpha \in G$ fixing the point 1 in X. If α commutes with γ then

$$1\gamma = 1\alpha\gamma = 1\gamma\alpha.$$

Hence α fixes two points, and so it is the identity. Therefore $C_G(\gamma) \cap G_i = \langle e \rangle$, for any point i in X. The element γ has $|G : C_G(\gamma)|$ distinct conjugates in G and these all have no fixed points. Hence $|G : C_G(\gamma)| \leq n - 1$ and $|C_G(\gamma)| \geq n$. If $|C_G(\gamma)| > n$ then $C_G(\gamma)$ must contain two elements sending 1 to the same point of X and therefore it contains a non-identity element fixing 1. Therefore $|C_G(\gamma)| = n$. It follows that $C_G(\gamma)$ consists of the identity and the $n - 1$ elements of G with no fixed points. It is therefore a regular normal subgroup of G. It also follows that the non-identity elements of $C_G(\gamma)$ are conjugate to γ, from which we deduce that they all have order p and that they commute. Thus $C_G(\gamma)$ is an elementary abelian p-group. □

It is well known, in some circles, that every sharply 2-transitive group of permutations corresponds to an affine plane (a so-called nearfield plane). If $\mathbb{F}$ is a field then the mappings $\tau_{a,b}$ defined on $\mathbb{F}$ by

$$\tau_{a,b}(x) = ax + b, \quad a \neq 0,$$

form a sharply 2-transitive group of permutations of $\mathbb{F}$.

A *Frobenius group* is a transitive permutation group such that only the identity fixes two or more elements, and some non-identity element fixes a point. Thus any sharply 2-transitive group is a Frobenius group. In a Frobenius group any two distinct point-stabilizers must intersect in the identity subgroup. Hence if G is a Frobenius group of degree n then it contains exactly $n-1$ elements with no fixed points. (For if a point-stabilizer has order m then $|G| = nm$ and there are exactly $n(m-1)$ elements fixing a single point, and the identity fixes n.) The dihedral groups of order $2p$ acting on a set of p sympols are Frobenius groups when p is prime. In ??Lemma 1.1 above we saw that the elements of G having no fixed points, together with the identity, formed a regular normal subgroup. This is true in any Frobenius group, but has only been proved using character theory. (The difficulty is to show that this set is a subgroup, and not just a subset.) This normal subgroup is known as the *kernel* of the Frobenius group. It was proved by Thompson that the kernel of a Frobenius group is nilpotent, i.e., its Sylow p-subgroups are normal for each p. From this it follows that an abstract group can have at most one representation as a Frobenius group. The existence of the kernel is easily established for Frobenius groups where the point-stabilizers have even order.

29.1.2 Lemma. *Let G be a Frobenius group acting on the set X, containing an element of even order which fixes a point. Then the elements of G with*

no fixed point, together with the identity form a regular abelian normal subgroup of G.

Proof. We may assume that $|X| = n$ and that G_1 has r involutions in it. Each of these fixes exactly one point, and hence is a product of $(n-1)/2$ disjoint transpositions. Then G contains rn involutions, which together use up $rn(n-1)/2$ transpositions. Since $\mathrm{Sym}(X)$ contains exactly $\binom{n}{2}$ transpositions, if $r > 1$ there must be two distinct involutions in G containing a common transposition. The product of these two elements thus fixes at least two points, and is therefore the identity. Thus we have shown that G_1 contains a unique involution.

Our next step is to show that the product of two distinct involutions has no fixed points. Let a and b be distinct involutions. If $i \in X$ and $iab = i$ then $ia = iabb = ib$. Hence both either $i = ia = ib$ or a and b contain the transposition (i, ia). The latter is impossible while if a and b both fix i then they are both involutions in G_i, hence they are equal.

Let T be the set of all involutions in G and let N be the set formed by the identity and the fixed-point free elements of G. By the previous paragraph, if $a \in T$ then $aT \subseteq N$. As T and N both have cardinality n, it follows that $aT = N$. Similarly $Ta = N$ and hence $TT = N$. Since $a^2 = e$,

$$NN = TaaT = TT = N$$

from we deduce that N is a subgroup of G. Since the set of fixed-point free elements of G is closed under conjugation, $N \trianglelefteq G$. Therefore N is regular normal subgroup of G. It remains to show that it is abelian. If $g \in N = aT$ then $g = ab$ for some b in T and $g^a = a^{-1}aba = ba = (ab)^{-1} = g^{-1}$. Hence if g and h lie in N then

$$g^{-1}h^{-1} = g^a h^a = (gh)^a = (gh)^{-1} = h^{-1}g^{-1}.$$

This shows that any two elements of N commute, and so N is abelian. □

Frobenius groups arise more often than might be expected. By way of example, we offer:

29.1.3 Lemma. *(Wielandt). Let G be a $\frac{3}{2}$-transitive permutation group on the set X. Then either G is primitive, or it is a Frobenius group.*

Proof. Suppose that G is $\frac{3}{2}$-transitive and imprimitive. Let B be set of imprimitivity for G with cardinality k and let the length of an orbit of G_1

on $X \setminus 1$ be m. Assume that $1 \in B$ and that $i \in X \setminus B$. Let $G_{i,B}$ be the subgroup of G_i fixing B as a set. We break up the proof into a number of steps.

29.1.4 Claim. *(a) k and m are coprime.*

For G_1 fixes B as a set, and so B consists of 1 and a number of non-trivial orbits of G_1. Hence m divides $k-1$.

29.1.5 Claim. *(b) $|G_i : G_{i,B}| = m$.*

As $i \notin B$ the set BG_i is both a disjoint union of orbits of G_i with length m, and a disjoint union of translates of B. Hence both k and m divide $N = |BG_i|$ and so km must divide N. Since each point of BG_i belongs to a G_i orbit of a point in B we have $N \leq km$, whence $N = km$. The number of translates of B in BG_i is equal to $|G_i : G_{i,B}|$, and therefore $|G_i : G_{i,B}| = m$.

29.1.6 Claim. *(c) Let b' be a point in $B \setminus b$. Then $G_{i,B} = G_{i,b} = G_{i,(B)} \leq G_{(B)} \leq G_{b,b'}$.*

(Here $G_{(B)}$ is the subgroup of G fixing each point in B and $G_{i,(B)} = G_i \cap G_{(B)}$.) For any point b in $X \setminus i$ we have that $m = |G_i : G_{i,b}|$. From this we deduce that if $b \in B$ and $i \notin B$ then $G_{i,b}$ and $G_{i,B}$ have the same index in G_i, and are consequently equal. In particular $G_{i,b}$ is independent of the choice of b in B, implying that $G_{i,b} = G_{i,(B)}$. Finally $G_{i,(B)} \leq G_{(B)}$ and $G_{(B)} \leq G_{b,b'}$.

29.1.7 Claim. *(d) $G_{i,b} = G_{b,b'} = G_{(B)}$.*

Since $G_{b,b'}$ has index m in G_b, it follows that $G_{b,b'}$ and $G_{i,b}$ have the same order, and so by (c) they are equal to each other and to $G_{(B)}$.

29.1.8 Claim. *(e) Let C be the unique translate of B containing i. Then $G_{i,b} = G_{(C)}$.*

Clearly $G_{b,i} = G_{i,b}$. If we replace b by i and B by C then we may repeat the arguments above to deduce that $G_{b,i} = G_{(C)}$.

We can now complete the proof of the lemma. From (d) and (e) we see that $G_{(B)} = G_{(C)}$, and so any element of G which fixes each point in B must fix each point in C. It follows that $G_{(B)} = G_{(Bg)}$ for any element g of G. Hence $G_{(B)} = \langle e \rangle$ and so, by (d), all stabilizers of two points are trivial. □

29.2 Normal Subgroups of 2-Transitive Groups

As the orbits of a normal subgroup form a system of imprimitivity and, as 2-transitive groups are necessarily primitive, a normal subgroup of 2-transitive group must be transitive. If $N \trianglelefteq G$ then $N_1 = N \cap G_1 \trianglelefteq G_1$, and therefore N_1 is a normal subgroup of G_1, which is transitive on the set of points other than 1. Ths shows that a normal subgroup of a 2-transitive group is $\frac{3}{2}$-transitive. But more can be said.

29.2.1 Lemma. *Let G be a 2-transitive permutation group on X and let N be a regular normal subgroup of G. Then N is an elementary abelian p-group. If the stabilizer of a point is primitive in its action on the remaining points then $p = 2$ or $|X| \leq 3$.*

Proof. Suppose that $g \in G_1$ and $h \in N$. Then

$$1hg = 1gg^{-1}hg = 1h^g.$$

As G_1 acts transitively on $X\backslash 1$, every point in $X\backslash 1$ can be written as $1h^g$ for some element g of G_1. As N is transitive $1h^g = 1k$ and as $N_1 = \langle e \rangle$, we thus have $h^g = k$. Thus any two non-identity elements of H are conjugate as elements of G, and so all elements of $H\backslash e$ have the same order. Hence H is a p-group. Since the centre of H is a characteristic subgroup of N, it is normal in G and so, if it is not transitive, its orbits form a system of imprimitivity for G. However G is 2-transitive, and hence primitive. Consequently $Z(H)$ is transitive. As $Z(H)_1 \trianglelefteq Z(H)$, it follows that $Z(H)$ must be regular. Hence H and $Z(H)$ have the same order, which means they are equal, and H is abelian. Thus we have shown that H is an elementary abelian p-group.

Suppose $p > 2$ and let h be an element of H. If $g \in G_1$ and $1h = 1hg$ then

$$1h = 1hg = 1gg^{-1}hg = 1h^g,$$

implying that $h = h^g$. It follows that $h^n = (h^n)^g$ for any integer n and hence that g fixes each point in $1\langle h \rangle$. Therefore $G_{1,1h}$ fixes at least $p - 1$ points. By Witt's theorem (??Theorem 1.7.4) the fixed points of $G_{1,1h}$ form a set of imprimitivity for G_1 in its action on $X \setminus 1$. So if $p > 2$ and G_1 is primitive then this set of imprimitivity must be trivial. Thus $G_{1,1h}$ must fix each point in $X \setminus 1$ and every element of N must be a power of h. Hence

N is a cyclic group of prime order and G_1 is a primitive regular group of degree $p-1$. It is easy to see that a primitive regular group must have prime degree. (Use the fact that the identity must be a maximal subgroup.) But if $p-1$ is a prime then $p=3$. □

Note that Sym(3) is 3-transitive and has a regular normal subgroup. Also Sym(4) is 4-transitive, and has a regular normal subgroup of order four. However this is the only 4-transitive group with a regular normal subgroup, and no 5-transitive group has a regular normal subgroup.

It is difficult to study 2-transitive groups without becoming involved with simple groups. Before proving one result which supports this assertion, some preliminaries are in order. The *commutator* of two subgroups H and K of a group G is defined to be

$$[H,K] := \langle h^{-1}k^{-1}hk : h \in H, k \in K \rangle.$$

We observe that $[H,K]$ is the identity subgroup if and only if the elements of H commute with the elements of K.

29.2.2 Lemma. *Let H and K be normal subgroups of the group G. Then $[H,K] \leq H \cap K$.*

Proof. If $H \trianglelefteq G$ and $h \in H$ then $k^{-1}hk \in H$ for all k in K. Hence $[H,K] \leq H$ in this case. If we also have that $K \trianglelefteq G$ then $[H,K] \leq K$, whence the lemma follows. □

29.2.3 Theorem. *(Burnside). Let N be a minimal normal subgroup of the 2-transitive group G. If N is not regular, it is primitive and simple.*

Proof. If $N \trianglelefteq G$ then it is $\frac{3}{2}$-transitive and therefore either primitive or Frobenius. If it is a Frobenius group then its kernel is a proper normal subgroup of G, hence N must be primitive.

Suppose that M is a minimal proper normal subgroup of N. Since N is primitive, M is transitive. As M is not normal in G, there is an element g of G such that $M \neq M^g$. Since $M \trianglelefteq N$, it follows that $M^g \trianglelefteq N$ and so $M \cap M^g \trianglelefteq N$. As M is minimal and not equal to M^g, this implies that $M \cap M^g = \langle e \rangle$. By ??? we now find that $[M, M^g] = \langle e \rangle$, and therefore the elements of M commute with the elements of M^g. If $x \in M$ and $1x = 1$ then, for any y in M^g,

$$1yx = 1xy = 1y$$

and so each point in $1M^g$ is fixed by x. Therefore M must be regular.

Suppose that $h \in G$ such that M^h is distinct from M and M^g and let H be the subgroup $\langle M, M^g \rangle$ of N. Then M^h and H are both normal in N, so $M^h \cap H \trianglelefteq N$. As M^h is a minimal normal subgroup of N, we deduce that either $M^h \cap H = \langle e \rangle$ or $M^h \leq H$. If the former holds then arguing as above we find that the elements of M^h and H commute, and hence that H is regular. This is impossible, since $|H| = |M|^2$ and M is regular. Therefore all subgroups of G conjugate to M are contained in H. Accordingly $H \trianglelefteq G$ and so, by the minimality of N we obtain that $N = H$. If M^h is distinct from M^g then $M^h M^g \leq H$ and

$$|M^h M^g| = |M^h||M^g|/|M^h \cap M^g| = |M|^2 = |H|,$$

whence $M^h M^g = H$. Further, an element of M commutes with any element of M^h or M^g, implying that $M \leq Z(H)$. Hence H is abelian, and is therefore regular. This is still impossible.

Thus we conclude that there are exactly two distinct conjugates of M in G. Consequently the normalizer K of N in G must have index two in G. As K contains N, it is transitive and so $G = G_1 K$. Since $K \trianglelefteq G$, it follows that $K_1 = K \cap G_1 \trianglelefteq G_1$. Furthermore,

$$|G : K| = |G_1 K : K| = \frac{|G_1 K|}{|K|} = \frac{|G_1||K|}{|G_1 \cap K||K|} = |G_1 : K_1|$$

and thus K_1 has index two in G_1.

Since $K_1 \trianglelefteq G_1$ its orbits form a system of imprimitivity for G_1 in its action on $X \setminus 1$. Hence G_1 induces a permutation group on the orbits of K_1. The stabilizer of an orbit in this action contains K_1, and so by the orbit-stabilizer relation, the number of orbits is at most $|G_1 : K_1| = 2$. If K_1 is transitive on $X \setminus 1$ then K is 2-transitive on X and M, as a normal subgroup of K, is abelian. Thus we may assume that K_1 has two orbits on $X \setminus 1$. Note that these orbits have same length, since they are permuted transitively by G_1.

The argument in the first part of the proof of ??stab1 now shows that the elements of M divide into two conjugacy classes, of equal size, under the action of K. Therefore the order of M is divisible by at most two distinct primes. If it is divisible by only one prime then it is a p-group and its centre is non-trivial and, since all elements of M are conjugate under the action of K_1, we deduce that M is abelian. Thus we may assume that there are

primes p and q such that every non-identity element of M has order p or q. Note that p and q are the only prime divisors of $m = |M|$. Let x be an element of order p lying in the centre of a Sylow p-subgroup P of M. Then $C_M(x)$ contains P and therefore $|M : C_M(x)| = q^r$ for some integer r. Since each conjugacy class of elements of M under the action of K is the disjoint union of conjugacy classes of M, the number of elements in M with order p is divisible by q. The two non-trivial orbits of K j=have length $(m-1)/2$; hence the number of p-elements in M is $(m-1)/2$ and thus q divides $m-1$. It also divides m which is coprime to $m-1$. This contradiction forces us to conclude that N has no proper normal subgroups, i.e., it is simple. □

This result indicates a close connection between 2-transitive permutation groups and the classification of the finite simple groups. Indeed, following the classification of the latter, it is possible to write down an explicit list of the 2-transitive permutation groups.

29.3 Sharply t-Transitive Groups

All the sharply t-transitive groups with $t \geq 2$ were determined long before the finite simple groups were classified. It is not too difficult to describe all sharply t-transitive groups with $t \geq 4$, as we will see. Three preliminary results are needed.

29.3.1 Lemma. *Let A be an abelian group of odd order and let τ be an automorphism of A with order two. Let $C_A(\tau)$ be the set of elements of A fixed by τ and let $I_A(\tau)$ be the elements mapped to their inverses by τ. Then $A = C_A(\tau)I_A(\tau)$ and $C_A(\tau) \cap I_A(\tau) = \langle e \rangle$.*

Proof. Since A has no elements of order two, $C_A(\tau) \cap I_A(\tau) = \langle e \rangle$. If $a \in A$ then

$$(aa^\tau)^\tau = a^\tau a = aa^\tau \in C_A(\tau)$$

and

$$(a(a^\tau)^{-1})^\tau = a^\tau a^{-1} = (a(a^\tau)^{-1})^{-1} \in I_A(\tau).$$

Further $(aa^\tau)a(a^\tau)^{-1} = a^2$. This shows that every element of A which is a square is the product of an element of $C_A(\tau)$ with an element of $I_A(\tau)$. As A has odd order the mapping $a :\mapsto a^2$ is onto; every element is a square. This proves the lemma. □

If $a = c_1 g_1 = c_2 g_2$, where $c_i \in C_A(\tau)$ and $g_i \in I_A(\tau)$ (for $i = 1, 2$) then

$$c_2^{-1} c_1 = g_2 g_1^{-1}.$$

Here the left side belongs to $C_A(\tau)$ while the right side is in $I_A(\tau)$. Hence both sides must equal the identity. This shows that an element of A can be expressed in at most one way as a product of an element of $C_A(\tau)$ and $I_A(\tau)$.

29.3.2 Lemma. *Let A be an elementary abelian 2-group and let τ be an automorphism of it with order two. Then $|A| \leq |C_A(\tau)|^2$.*

Proof. Let B be the set

$$\{a^{-1} a^\tau : a \in A\}.$$

We note that B is a subgroup of A and the mapping

$$\tau - 1 : a \mapsto a^{-1} a^\tau$$

is a homomorphism of A onto B. Hence

$$|B| = |A| / \ker(\tau - 1).$$

But $B \leq C_A(\tau)$, since $a = a^{-1}$ for every element a of A. Further, $C_A(\tau) = \ker(\tau - 1)$, whence we obtain that $|C_A(\tau)| \geq |A|/|C_A(\tau)|$. □

Our next result is another transitivity lemma.

29.3.3 Lemma. *Let G be a group acting t-transitively on X. If Q is a p-subgroup of G, maximal subject to fixing at least $t + 1$ points, then $N_G(Q)$ acts t-transitively on* $\operatorname{fix} Q$.

Proof. Let T be a t-subset of X and let $G_{(T)}$ denote the subgroup of T fixing each point in T. We may assume without loss that $Q \leq G_{(T)}$. If Q is a Sylow p-subgroup of $G_{(T)}$ then the claim follows from Witt's theorem (??Theorem 1.7.4). Otherwise let S be set of t points from $\operatorname{fix} Q$. Then Q is a proper subgroup of a Sylow p-subgroup P of $G_{(S)}$. From our assumption on Q we see that $S = \operatorname{fix} P$. Hence each t-subset of $\operatorname{fix} Q$ is the fixed point set of the p-group $N_P(Q)$. As $Q \trianglelefteq N_G(Q)$, its fixed point set is a union of orbits of $N_G(Q)$. By the t-transitive version of Gleason's lemma, it follows that $N_G(Q)$ is t-transitive on $\operatorname{fix} Q$. □

29.3.4 Lemma. *Let G be a 3-transitive group on X. If the stabilizer of any three points has even order and the stabilizer of any four is trivial then $|X| = 5$ or 11, and G is sharply 4-transitive.*

Proof. Let x and y be two points of X and set H equal to G_{xy}. Then H is a Frobenius group on $X' = X \setminus xy$. By hypothesis the point stabilizers in H have even order and thus, by ??frob, we find that H has a regular abelian normal subgroup A. If $z \in X'$ then the proof of ??frob shows that H_z contains a unique involution t and that $a^t = a^{-1}$ for each element a of A. If $a \in A$ and $zag = za$ then

$$za = ztt^{-1}at = za^t.$$

Then $a^t \in A$ and $a^t a^{-1}$ fixes z. Since A is regular on $X \setminus xy$, this implies that $a^t = a$. Thus there is a bijection between the points of $X \setminus xy$ fixed by t and the elements of A fixed by t.

Since t is an involution, it interchanges some pair of elements, u and v say, from X'. As G_z is 2-transitive, there is an element g in G mapping u and v respectively to x and y. Let $s = t^g$. Then s fixes z and interchanges x and y. Hence it fixes X' as a set. From this it follows that if $a \in A$ then a^s is a permutation fixing X' as a set and acting fixed-point freely on it. Therefore a^s must lie in the kernel of the Frobenius group H, i.e., $a^S \in A$. Thus s normalizes A. Since t has exactly three fixed points, so does s. Hence $|C_A(s)| = 3$. The subgroup H_z is also normalized by s, from which it follows that $t^s = t$, since t is the unique involution in H_z. Thus t and s commute and so $ts = st$ is an involution. Since the stabilizer of any four points of X is trivial, ts must fix one or three points. Hence $|C_A(gs)| \leq 3$. As an element of A is fixed by ts if and only if it is inverted by s, we thus have $|I_A(s)| = 1$ or 3.

From ??ab1 we now deduce that $|A| = 3$ or 9, and therefore $|X| = 5$ or 11. If $|X| = 5$ then $|G| = 5 \cdot 4 \cdot 3 \cdot 2$ and G is sharply 4-transitive. Suppose that $|X| = 11$. Then

$$|G| = 11 \cdot 10 \cdot 9 \cdot k,$$

where $k = 2$, 4, or 8. Let P be a Sylow 11-subgroup of G. We need information about the order of $N = N_G(P)$. It is a transitive group on X, with P as a regular normal subgroup. Hence $|N| = |N_1||P|$. The only element of N_1 which commutes with each element of P is the identity. Hence, if $h \in N$ then the mapping $x \mapsto x^h$ is a non-identity automorphism of P.

Thus N_1 is isomorphic to a subgroup of $\mathrm{Aut}(P)$ and, as $|\mathrm{Aut}(P)| = 10$, we deduce that $|N_1|$ divides 10. Therefore $|G : N_G(P)|$ is equal to $9k$ times a divisor, ℓ say, of 10. Since $|G : N_G(P)|$ equals the number of Sylow 11-subgroups of G, it must also be congruent to 1 modulo 11. Thus we are forced to conclude that $\ell = 2$ and $k = 8$, and hence that G is sharply 4-transitive. □

29.3.5 Lemma. *If G is a sharply 4-transitive group on X then $|X| = 4$, 5, 6 or 11.*

Proof. If $|X|$ is odd then we appeal to the previous lemma. Suppose that $|X|$ is even and that x, y and z are distinct elements of X. Then G_{xy} is a sharply 2-transitive group, on $X' = X \setminus xy$. Hence it contains an elementary abelian normal subgroup A, and G_z contains a unique involution t such that $a^t = a^{-1}$ for all elements a of A. Note that since $|X'|$ is even, A is 2-group. Since t fixes at most two points of X' we see that $|C_A(g)| \leq 2$ and so, by Lemma 3.2, $|A| \leq 4$. Hence $|X| \leq 6$. □

29.4 Generously k-Transitive Groups

We call a group G on X *generously k-transitive* if the stabilizer of any $k-1$ points is generously transitive on the remaining points. If $Y \subseteq X$, let G_Y be the subgroup of G fixing Y as a set.

29.4.1 Lemma. *A group G is generously k-transitive on X if and only $G_Y \restriction Y$ is isomorphic to $\mathrm{Sym}(Y)$, for any subset Y in $\binom{X}{k+1}$.*

Proof. Suppose G is generously k-transitive and $Y \in \binom{X}{k+1}$. If x and y are elements of Y then there is an element τ of G that fixes each element of $Y \setminus \{x, y\}$ and swaps x and y. Now $\tau \restriction Y$ is a transposition, and so we deduce that $G_Y \restriction Y$ contains all transpositions of the elements of Y. Hence it is isomorphic to $\mathrm{Sym}(Y)$. The converse is immediate. □

29.4.2 Lemma. *Let G be a generously k-transitive group on X and let Y be a k-subset of X. Then $G_{(Y)}$ and G_Y have the same orbits on $X \setminus Y$.*

Proof. Clearly $G_{(Y)} \trianglelefteq G_Y$. Suppose that a and b are elements of $X \setminus Y$ and $g \in G_Y$ such that $ag = b$. Since the restriction of $G_{Y \cup a}$ to $Y \cup a$ is

the symmetric group, there is an element h of G such that $ah = a$ and $h\restriction Y = g\restriction Y$. Then $ahg = ag = b$ and $hg\restriction Y$ is the identity. □

29.4.3 Lemma. *(Wagner). Let G be a 3-transitive group on X, where $|X|$ is odd and greater than three. Then any non-identity normal subgroup of G is 3-transitive.*

Proof. Suppose N is a non-identity normal subgroup of G. If N is regular then, by ??Lemma 2.1, either $|N| = 3$ or N is a 2-group. Therefore N is not regular, and hence it is $\frac{5}{2}$-transitive. In particular $|N|$ is even. Let t be an involution in N and let $\Delta = \{1, 2, 3\}$ be a subset of X. As G is 3-transitive, we may assume that $1t = 1$ and $2t = 3$. Thus

$$(12) \in (N_\Delta\restriction\Delta) \trianglelefteq (G_\Delta\restriction\Delta) \cong \mathrm{Sym}(3).$$

The only normal subgroup of Sym(3) containing a transposition is Sym(3) itself. It follows now that N must be generously 2-transitive. As N is $\frac{5}{2}$-transitive, N_{12} is $\frac{1}{2}$-transitive on $X \setminus \{1, 2\}$. Hence all orbits of N_{12} on $X\setminus\{1, 2\}$ have length dividing $|X|-2$, which is odd (in the number-theoretic sense). Suppose N is not 3-transitive. Then N_{12} must have at least two odd orbits on $X \setminus \{1, 2\}$. Let P be a Sylow 2-subgroup of $M = N_{\{1,2\}}$. Since M and N_{12} have the same orbits on $X \setminus \{1, 2\}$ it follows that M has two odd orbits on $X \setminus \{1, 2\}$ and so P must fix at least two points. Consequently there is an element g of G such that $P^g \leq N_{12}$. This implies that $|M : N_{12}|$ is odd, which is odd (even impossible) since $|M : N_{12}| = 2$. □

This result shows that a sharply 3-transitive group of odd degree must be simple. If $q = 2^k$ then the group $PGL(2, q)$ acts sharply 3-transitively on a set of size $q + 1$ and is thus simple when $q > 2$. In particular we find that Alt(5) is simple.

Exercises

1. If G is a $\frac{5}{2}$-transitive group with a regular normal subgroup, show that it has degree four. Show that a 5-transitive group cannot have a regular normal subgroup.

2. If G is 2-transitive on X and x and y are distinct points from X, show that $G_{\{x,y\}}$ has at most one orbit of odd length.

3. If G acts sharply 5-transitively on V, show that $|V| = 11$.

Chapter 30

Association Schemes

30.1 Definition

First an example. Let $\mathcal{D}$ be a design with blocks of size k, and define matrices $A_0, \ldots, A_d$ with rows and columns indexed by the blocks of $\mathcal{D}$, such that

$$(A_i)_{\alpha,\beta} = \begin{cases} 1, & |\alpha \cap \beta| = k - r; \\ 0, & \text{otherwise.} \end{cases}$$

So the A_i are symmetric 01-matrices that sum to J, and $A_0 = I$. We call them the *block intersection matrices* of the design. Any automorphism of $\mathcal{D}$ must commute with each of the matrices A_i, and hence it must commute with any matrix in the algebra that they generate. Thus in some sense our design is most regular if the dimension of this algebra is as small as possible. Since $A_0, \ldots, A_d$ are linearly independent, the lower bound on this dimension is one ore than the degree of $\mathcal{D}$, and if equality holds then there are interesting consequences.

First, for each i and j the product $A_i A_j$ belongs to the algebra, and therefore there must be scalars $p_{i,j}(r)$ such that

$$A_i A_j = \sum_{r=0}^{k} p_{i,j}(r) A_r.$$

Second, from this we see that $A_i A_j$ is symmetric, which implies that A_i and A_j commute.

The opreceding discussion may motivate the following definijtion. An *association scheme* with d classes is a set $\mathcal{A}$ of 01-matrices $A_0, \ldots, A_d$ such that

(a) $A_0 = I$.

(b) $\sum_i A_i = J$.

(c) $A_i^T \in \mathcal{A}$ for each i.

(d) There are scalars $p_{i,j}(r)$ such that $A_i A_j = \sum_i p_{i,j}(r) A_r$.

(e) $A_i A_j = A_j A_i$ for all i and j.

If $A_i = A_i^T$, we say that the scheme is *symmetric*. This is the only case we will consider (and then (e) is redundant). The matrices $A_1, \ldots, A_d$ can be viewed as adjacency matrices of graphs $X_1, \ldots, X_d$. We will say that these graphs form an association scheme—this means an association scheme is a set of matrices or a set of graphs, whichever suits us. We use $\mathbb{R}[\mathcal{A}]$ to denote the vector space spanned by the matrices in the scheme; this is known as its *Bose-Mesner algebra*. Since $J \in \mathbb{R}[\mathcal{A}]$ and since $\mathbb{R}[\mathcal{A}]$ is commutative, each matrix A_i commutes with J and therefore each graph in the scheme is regular.

30.2 Schematic Designs

30.2.1 Theorem. *Let $\mathcal{D}$ be a design with degree s and strength t. If $t \geq 2s - 2$, then the block intersection matrices of $\mathcal{D}$ from an association scheme with s classes.*

A design is said to be *schematic* if its intersection matrices form an association scheme. It is an exception for a design to be schematic. Any symmetric design is trivially schematic. A 2-design with degree two is schematic, and so each 2-$(v, k, 1)$ gives us an association scheme with two classes. A graph X and its complement form an association scheme with two classes if and only if X (and $\overline{X}$) are strongly regular.

Using Hamming distance we can define "intersection matrices" for an orthogonal array. The analog of Theorem 30.2.1 holds for orthogonal arrays as well. If an orthogonal array has degree s and strength t and $t \geq 2s - 2$, then its intersection matrices form an association scheme with s classes.

Accordingly each orthogonal array with index one and strength two gives rise to a pair of strongly regular graphs.

A finite set S of points on the unit sphere in $\mathbb{R}^d$ has strength at least t if the average over the points in the set of a polynomial of degree at most t is equal to its average over the entire sphere. Then S is a 2-design if and only if

$$\sum_{x \in S} xx^T = \frac{|S|}{d} I.$$

The degree of S is the size of the set of inner products $x^T y$, for distinct points x and y. If S has degree s and strength t and $t \geq 2s - 2$, we get an association scheme with s classes.

Chapter 31

Matrix Theory

31.1 The Kronecker Product

If A and B are matrices over the same ring, we define their *Kronecker product* $A \otimes B$ to be the matrix we get by replacing each entry $A_{i,j}$ of A with the matrix $A_{i,j}B$. Note that neither A nor B need be square. For example, if $x \in \mathbb{F}^m$ and $y \in \mathbb{F}^n$, then

$$x \otimes y^T = xy^T.$$

If A is an $m \times n$ matrix, then $\operatorname{vec}(A)$ is the $mn \times 1$ matrix we get by stacking the columns of A one on top of the other. So if $e_1, \ldots, e_n$ is the standard basis, then

$$\operatorname{vec}(A) = \begin{pmatrix} Ae_1 \\ \vdots \\ Ae_n \end{pmatrix}$$

The Kronecker product is bilinear, i.e., it is linear in each variable. We also have

$$(A \otimes B)^T = A^T \otimes B^T.$$

The following properties are fundamental.

31.1.1 Theorem. *If A, X and B are matrices such that the product AXB^T is defined, then*

$$(I \otimes A)\operatorname{vec}(X) = \operatorname{vec}(AX), \qquad (B \otimes I)\operatorname{vec}(X) = \operatorname{vec}(XB^T). \qquad \square$$

One consequence of this is that

$$(A \times B) = (I \otimes A)(B \otimes I) = (B \otimes I)(I \otimes A).$$

It also follows that if AC and BD are defined, then

$$(A \otimes B)(C \times D) = AC \otimes BD.$$

In particular if $Ax = \lambda x$ and $By = \mu y$, then

$$(A \otimes B)(x \otimes y) = \lambda x \otimes \mu y = \lambda\mu x \otimes y.$$

We have

$$(e_i \otimes g_k)^T (A \otimes B)(f_j \otimes h_\ell) = e_i^T A f_j \, g_k^T B h_\ell.$$

Because of this we can view the rows of $A \times B$ as being indexed by ordered pairs (i, k), and the columns by ordered pairs (j, ℓ). Then

$$(A \otimes B)_{((i,k),(j,\ell))} = A_{i,j} B_{k,\ell}$$

Suppose U and V are vector spaces. The vector space spanned by the vectors

$$u \otimes v, \quad u \in U, \ v \in V$$

is called the *tensor product* of U and V, and is denoted by $U \otimes V$. (I will become upset if refer to this as the Kronecker product of U and V.)

Suppose P is the linear mapping on $V \otimes V$ defined by the requirement that

$$P(x \times y) = y \otimes x.$$

Prove that $P^2 = I$, that P commutes with $A \otimes A^T$ and that $P(A \otimes A^T)$ is symmetric

If A and B are $m \times n$ matrices, we define their *Schur product* $A \circ B$ by

$$(A \circ B)_{i,j} := A_{i,j} B_{i,j}$$

(It is sometimes called the bad-student's product.) Show that $A \circ B$ is a principal submatrix of $A \otimes B$.

31.2 Normal Matrices

A matrix M over $\mathbb{C}$ is *normal* if it commutes with its conjugate-transpose. Examples are Hermitian matrices and unitary matrices. If $A = L^*DL$ where L is unitary and D is diagonal, then $A^* = L^*\overline{D}L$; hence A is normal. So a matrix that unitarily diagonalizable is normal.

The converse is true:

31.2.1 Theorem. *A matrix M is unitarily diagonalizable if and only if it is normal.* □

31.2.2 Theorem. *Suppose $\mathcal{A}$ is a commutative algebra of $v \times v$ complex matrices. If $\mathcal{A}$ is closed under complex-conjugate, then their is a basis for $\mathbb{C}^v$ that consists of common eigenvectors for (the matrices in) $\mathcal{A}$.* □

This theorem fails for the commutative algebra consisting of the matrices of the form

$$\begin{pmatrix} a & b \\ 0 & a \end{pmatrix}, \quad a, b \in \mathbb{R}.$$

A square matrix N is normal if and only if for all vectors z we have

$$\langle Nz, Nz\rangle = \langle N^*z, N^*z\rangle. \tag{31.2.1}$$

It is easy to verify that this holds if N is normal:

$$\langle Nz, Nz\rangle = z^*N^*Nz = z^*NN^*z = \langle N^*z, N^*z\rangle.$$

For the converse, show that (31.2.1) holds if and only if

$$\langle Nz, Nw\rangle = \langle N^*z, N^*w\rangle$$

for all z and w, and then show that $z^*(N^*N - NN^*)w = 0$ for all z and w if and only if $N^*N = NN^*$.

31.3 Positive Semidefinite Matrices

A complex matrix M is *positive semidefinite* if $M = M^*$ and $z^*Mz \geq 0$ for all z. If M is positive semidefinite and $z^Mz = 0$ implies that $z = 0$, it is *positive definite*.

31.3.1 Theorem. *If M is a Hermitian matrix, the following assertions are equivalent:*

(a) M is positive semidefinite.

(b) $M = N^*N$ for some matrix N.

(c) The eigenvalues of M are non-negative. □

A positive semidefinite matrix is positive definite if and only if it is invertible. The sum of two positive semidefinite matrices is positive semidefinite, and the sum of a positive definite and a positive semidefinite matrix is positive definite, hence invertible. Thus if $r > \lambda$ and $NN^T = (r-\lambda)I + \lambda J$, then NN^T is invertible because $(r-\lambda)I$ is positive definite and $\lambda J = \lambda \mathbf{1}\mathbf{1}^T$ is positive semidefinite.

31.3.2 Lemma. *A Hermitian matrix M is positive semidefinite if and only* $\mathrm{tr}(MX) \geq 0$ *for all positive semidefinite matrices X.* □

(The set of $n \times n$ positive semidefinite matrices is a convex cone; this lemma implies that this cone is self dual. Talk to Levent Tuncel.)

Prove that a principal submatrix of a positive semidefinite matrix is positive semidefinite. This implies that the diagonal entries of a positive semidefinite matrix are non-negative. Prove that if M is positive semidefinite and $M_{r,r} = 0$, then all entries in the r-th row and all entries in the r-th column of M are zero.

Prove that if M and N are positive semidefinite, so is $M \otimes N$. Deduce that if M and N have the same order then $M \circ N$ is also positive semidefinite. (This is an important result due to Schur.)

31.4 Tight Partitions

A partition of V is a set whose elements, which we usually call *cells*, are subsets of V. The *characteristic matrix* of P is the matrix whose columns are the characteristic vectors of the cells of π. We use $|\pi|$ to denote the number of cells of π, and so the characteristic matrix of π has order $|V| \times |\pi|$. The columns of the characteristic matrix are linearly independent.

If π is a partition of the points of a finite incidence structure. Declare blocks α and β of the structure to be π-equivalent if, for each cell C of π,

$$|C \cap \alpha| = |C \cap \beta|.$$

The π-equivalence classes are a partition of the blocks, which we call the *induced partition*. We consider one important example. Suppose G is a group of automorphisms of an incidence structure, let ρ be the partition formed by the orbits of G on points and let σ be the partition formed by the orbits of G on blocks. Then σ is a refinement of the induced partition ρ^*. (And thus $|\sigma| \geq |\rho^*|$.)

31.4.1 Theorem. *Suppose $\mathcal{S}$ is a finite incidence structure and the rows of the incidence matrix of $\mathcal{S}$ are linearly independent. If π is a partition of the points and π^* is the induced partition on blocks, then $|\pi| \leq |\pi^*|$.*

Proof. Let N be the incidence matrix of $\mathcal{S}$ and let P be the characteristic matrix of π. Two blocks are π-equivalent if and only if the corresponding columns of $P^T N$ are equal, and thus $|\pi^*|$ is the number of distinct columns of $P^T N$. Hence $\mathrm{rk}(P^T N) \leq |\pi^*|$.

We claim that the rows of $P^T N$ are linearly independent. For if $x^T P^T N = 0$ then, since the rows of N are linearly independent, $x^T P^T = 0$. However the rows of P^T are linearly independent and therefore $x = 0$. It follows that $\mathrm{rk}(P^T N) = |\pi|$, and the theorem follows. □

If the incidence matrix of an incidence structure has linearly independent rows and π is a partition of the points such that $|\pi| = |\pi^*|$, we say that π is a *tight partition*.

As is well known, the incidence matrix of a 2-design has linearly independent rows. We deduce that if G is a group of automorphisms of a 2-design then the number of orbits of G on blocks is at least as large as the number of orbits on points. As a corollary we see that if G acts transitively on blocks, it must also act transitively on points. The points and hyperplanes of a projective geometry form a 2-design, and we see that the number of orbits of a group G of collineations on hyperplanes is as least as large as the number of orbits on points. By duality we also get the reverse inequality, whence we have that G has equally many orbits on points and on hyperplanes.

The incidence matrix of a thick generalized quadrangle does not have linearly independent rows.

An incidence structure is *square* if its point and block sets have the same size.

31.4.2 Lemma. *If $\mathcal{S}$ is a square incidence structure whose incidence matrix is invertible, then any automorphism of $\mathcal{S}$ fixes equally many points and blocks.*

Proof. Let N be an incidence matrix of $\mathcal{S}$. An automorphism of $\mathcal{S}$ is a pair of permutation matrices (P, Q) such that $PNQ^T = N$. If N is invertible then

$$N^{-1}PN = Q$$

and therefore $\mathrm{tr}(P) = \mathrm{tr}(Q)$. Since $\mathrm{tr}(P)$ is the number of fixed points and $\mathrm{tr}(Q)$ the number of fixed blocks, this proves the lemma. □

Burnside's lemma informs us that the number of orbits of a permutation group on a set is equal to the average number of points fixed by a group element. So the lemma implies that, for a group of automorphisms of a square incidence structure with invertible incidence matrix, the number of orbits on points is equal to the number of orbits on blocks. It does **not** follow that the lengths of point orbits coincide with the length of the block orbits. Consider for example a projective plane over a field of order q and let G be the group of collineations that fix a point p. Then G has two orbits on points, of length 1 and $q^2 + q$, and two orbits on lines, of length $q + 1$ and q^2.

Suppose π is a partition of the points of an incidence structure with incidence matrix N and P is the characteristic matrix of π. Let Q be the characteristic matrix of the induced partition. If M is the matrix formed from the distinct columns of $P^T N$, then

$$P^T N = MR^T.$$

Here M is of order $|\pi| \times |\pi^*|$, and if the rows of N are linearly independent then so are the rows of M—if $x^T M = 0$ then $x^T P^T N = 0$. If the partition π is tight, then M is square and invertible. We call M the *quotient matrix* of N relative to π and π^*. (For the purposes of this discussion, we may replace π^* by any refinement and can still form a quotient matrix.)

A partition π^* on the blocks of an incidence structure induces a partition of its points, this will in general be a refinement of π.

Chapter 32

Finite Fields

32.1 Arithmetic

We now start to develop the theory of finite fields. Let $\mathbb{F}$ be a field. The identity element of $\mathbb{F}$ generates an additive subgroup of $\mathbb{F}$. If this subgroup is infinite, we say that $\mathbb{F}$ has *characteristic zero*. If $\mathbb{F}$ has characteristic zero, the subgroup generated by 1 is isomorphic to $\mathbb{Z}$; it follows that $\mathbb{F}$ contains a subfield isomorphic to $\mathbb{Q}$. This is the prime subfield of $\mathbb{F}$—the minimal subfield that contains 1. Both $\mathbb{R}$ and $\mathbb{C}$ are fields of characteristic zero.

Alternatively, the subgroup generated by 1 is finite. In this case the characteristic of $\mathbb{F}$ is defined to be the order of this subgroup. If p is prime and $\mathbb{F} = \mathbb{Z}_p$, then the characteristic of $\mathbb{F}$ is p. The field of rational functions over $\mathbb{Z}_p$ is infinite, but its characteristic is p.

32.1.1 Theorem. *Let $\mathbb{F}$ be a field. If the characteristic of $\mathbb{F}$ is not zero, it is a prime number.*

Proof. Suppose the characteristic of $\mathbb{F}$ is n and $n = ab$. Then $n \cdot 1 = 0$ and therefore

$$(a \cdot 1)(b \cdot 1) = n \cdot 1 = 0.$$

Since $\mathbb{F}$ is a field, this implies that either $a \cdot 1 = 0$ or $b \cdot = 1$. If $a \cdot 1 = 0$ then the order of the subgroup generated by 1 divides a. Therefore $a = n$ and $b = 1$. We conclude that n must be a prime number. □

If the characteristic of $\mathbb{F}$ is positive, it follows that the prime field of $\mathbb{F}$ is isomorphic to $\mathbb{Z}_p$, for some prime p. Hence $\mathbb{F}$ is a vector space over $\mathbb{Z}_p$

for some prime p, and consequently

$$|\mathbb{F}| = p^n$$

for some integer n.

32.1.2 Corollary. *The order of a finite field is the power of a prime.* □

We turn from addition to multiplication. The non-zero elements of a field form an abelian group under multiplication, when the field is finite we have a much stronger assertion.

32.1.3 Theorem. *If $\mathbb{F}$ is finite, its non-zero elements form a cyclic group under multiplication.*

Proof. Let A denote the group formed by the non-zero elements of $\mathbb{F}$. Let m denote the exponent of A, that is, the least integer m such that $a^m = 1$ for all a in A. Then every element of A is root of the polynomial $t^m - 1$. Since $t^m - 1$ has at most m roots, $|A| \leq m$. On the other hand, by Lemma **??**, we see that A contains an element of order m, and therefore it is cyclic. □

We use $\mathbb{F}^*$ to denote the set of non-zero elements of $\mathbb{F}$. A generator of the cyclic group $\mathbb{F}^*$ is usually called a *primitive element* of $\mathbb{F}$.

32.1.4 Corollary. *If $\mathbb{F}$ is a finite field of order q and $a \in \mathbb{F}$, the minimal polynomial of a divides $t^{q-1} - 1$.*

Proof. The non-identity elements of $\mathbb{F}$ form a cyclic group of order $q-1$, and therefore $a^{q-1} - 1 = 0$. So the minimal polynomial of a divides $t^{q-1} - 1$. □

32.2 Automorphisms

We study the automorphism of finite fields. Let $\mathbb{F}$ be a field. A map $\sigma : \mathbb{F} \to \mathbb{F}$ is an automorphism if it is a bijection and, for all a and b in $\mathbb{F}$,

$$(a + b)^\sigma = a^\sigma + b^\sigma, \quad (ab)^\sigma = a^\sigma b^\sigma.$$

The most familiar example is the operation of complex conjugation on the complex numbers. We describe a second example. Let $\mathbb{F}$ be a field with characteristic p. If i is an integer and $0 < i < p$ then the binomial coefficient

$$\binom{p}{i} = 0$$

modulo p. Hence if x and y belong to $\mathbb{F}$, then

$$(x+y)^p = x^p + y^p.$$

Since $(xy)^p = x^p y^p$, it follows that the p-th power map $x \mapsto x^p$ is an automorphism of $\mathbb{F}$. It is known as the *Frobenius automorphism* of $\mathbb{F}$.

If γ is an automorphism of the field $\mathbb{F}$, then $\mathrm{fix}(\gamma)$ is the subset

$$\{a \in \mathbb{F} : a^\gamma = a\}.$$

We say that $\mathrm{fix}(\gamma)$ is the set of elements of $\mathbb{F}$ fixed by γ. If Γ is a group of automorphisms of $\mathbb{F}$, then $\mathrm{fix}(\Gamma)$ denotes the set of elements of $\mathbb{F}$ fixed by each element of Γ. Hence

$$\mathrm{fix}(\Gamma) = \bigcap_{\gamma \in \Gamma} \mathrm{fix}(\gamma).$$

It is easy to verify that $\mathrm{fix}(\gamma)$ is a subfield of $\mathbb{F}$, and therefore $\mathrm{fix}(\Gamma)$ is a subsfield too. For example, if γ is complex conjugation on $\mathbb{C}$, then $\mathrm{fix}(\gamma) = \mathbb{R}$.

32.2.1 Lemma. *Let $\mathbb{F}$ be a field of characteristic p, and let τ be the Frobenius automorphism of $\mathbb{F}$. Then $\mathrm{fix}(\gamma)$ is the prime subfield of $\mathbb{F}$.*

Proof. We have $a^\tau = a$ if and only if $a^p - a = 0$. Therefore $\mathrm{fix}(\tau)$ is a subfield of $\mathbb{F}$ consisting of the roots of $t^p - t$, and therefore it is the prime subfield.

If $\gamma \in \mathrm{Aut}(\mathbb{F})$, then for any a in $\mathbb{F}$

$$a^\gamma = (1a)^\gamma = 1^\gamma a^\gamma$$

and so $1^\gamma = 1$. Now

$$(1+1)^\gamma = 1^\gamma + 1^\gamma = 1 + 1,$$

and a very simple induction argument yields that each element of the prime subfield is fixed by γ.

32.2.2 Theorem. *Let $\mathbb{F}$ be a finite field of characteristic p and order p^n and let τ be the Frobenius automorphism of $\mathbb{F}$. Then $\mathbb{F}$ has a subfield of order p^k if and only if $k \mid n$. If $k \mid n$, then there is a unique subfield of order p^k; it is the fixed field of τ^k.*

Proof. Let $\mathbb{K}$ be a subfield of $\mathbb{F}$ with order $q = p^k$. Then $\mathbb{F}$ is a vector space over $\mathbb{K}$, and therefore there is an integer d such that

$$p^n = |\mathbb{F}| = |\mathbb{K}|^d = p^{kd}.$$

This shows that $k \mid n$. Each element of $\mathbb{K}$ is a root of $t^{p^k} - t$, since this polynomial has at most p^k roots in $\mathbb{F}$, there is at most one field of order p^k.

If $a \in \mathbb{F}$ then $a^{\tau^k} = a$ if and only if

$$a^{p^k} - a = 0.$$

Accordingly $\mathrm{fix}(\tau^k)$ consists of roots of

$$t^{p^k} - t.$$

If $k \mid n$, then $t^{p^k} - t$ divides $t^q - t$ and therefore it has exactly p^k roots in $\mathbb{F}$. Thus $\mathrm{fix}(\tau^k)$ is a subfield of order p^k. □

32.2.3 Corollary. *If $\mathbb{F}$ is a finite field of characteristic p and order p^d, then the Frobenius automorphism of $\mathbb{F}$ has order d.* □

32.2.4 Theorem. *Let $q = p^d$, where p is prime. There is a unique field of order q, which is the splitting field for the polynomial $t^q - t$.*

Proof. Let $\mathbb{F}$ denote the splitting field for $t^q - t$ over the field $\mathbb{Z}_p$. Then the q roots of $t^q - t$ in $\mathbb{F}$ form the set of elements of $\mathbb{F}$ fixed by the q-th power map $a :\mapsto a^q$, and therefore they form a subfield of $\mathbb{F}$. Since this subfield contains all roots of $t^q - q$, it is the splitting field for $t^q - t$ and therefore this subfield equals $\mathbb{F}$. This shoiws that a field of order q exists.

Suppose $\mathbb{E}$ is a field of order q. Since $\mathbb{F}^*$ is cyclic, the elements of $\mathbb{F}$ give q distinct roots for $t^q - t$, Since no subfield of $\mathbb{E}$ can contain all these roots, $\mathbb{E}$ is a splitting field for $t^q - t$. As all splitting fields for a polynomial are isomorphic, this shows that there is a unique field of order q (up to isomorphism). □

Let q be a prime power and let $\mathbb{E}$ be an extension of degree d of a field $\mathbb{F}$ with order q. Let a be a primitive element in $\mathbb{E}$ and let ψ be its minimal polynomial over $\mathbb{F}$. Then ψ is irreducible over $\mathbb{F}$ and

$$\mathbb{E} = \mathbb{F}(a) \cong \mathbb{F}[t]/(\psi).$$

This implies that $\deg(\psi) = n$. By the theorem, for each positive integer d there is a finite field $\mathbb{E}$ of order q^d, and this field has a subfield of order q. We conclude that for each positive integer d, there is an irreducible polynomial in $\mathbb{F}[t]$ with degree d.

32.3 Squares

We describe the basic results concerning squares in finite fields.

An element in a field $\mathbb{F}$ that can be written as a^2 for some a is, naturally, called a *square* in $\mathbb{F}$. We determine the squares in finite fields.

First let $\mathbb{F}$ be a finite field of characteristic two. Then $\mathbb{F}^*$ is a cyclic group of odd order. Every element in a group of order order is square. For suppose x is a group element with odd order k. Then $k+1$ is even and so

$$x = \left(x^{(k+1)/2}\right)^2;$$

therefore x is square.

Now consider a field $\mathbb{F}$ of order q, where q is odd. Then $\mathbb{F}^*$ is a cyclic group of even order. Let a be a primitive element of $\mathbb{F}$, that is, a generator of $\mathbb{F}^*$. The non-zero squares in $\mathbb{F}$ are precisely the even powers of a, and these form a subgroup of $\mathbb{F}^*$ of order $(q-1)/2$. Thus exactly half the non-zero elements of $\mathbb{F}$ are squares.

Denote the set of non-zero squares in $\mathbb{F}$ by S. Since the index of S in $\mathbb{F}^*$ is two, the quotient $\mathbb{F}^*/S$ is isomorphic to the subgroup of $\mathbb{Z}$ formed by the set $\{1, -1\}$. Therefore map from $\mathbb{F}^*$ to $\mathbb{Z}$ that assigns 1 to each square and -1 to each non-square is a homomorphism. It follows that the product a non-zero square with a non-square is not a square, and the product of two non-squares is a square.

As a has order $q-1$, we see that $a^{(q-1)/2}$ is a root of $t^2 - 1$ and is not equal to 1. Hence $a^{(q-1)/2} = -1$ and from this we deduce that -1 is a square if and only if $(q-1)2$ is even. In other words, -1 is a square if and only if $q \equiv 1 \mod 4$.

Finally we prove that each element of $\mathbb{F}^*$ is the sum of two squares. Consider the set $S + S$. Since the order of S does not divide q, we see that S is not an additive subgroup of $\mathbb{F}$ and therefore there is an element b in $(S+S) \setminus S$. Since b is not a square, the multiplicative coset bS is the set of non-zero non-squares in $\mathbb{F}^*$ and, as b is the sum of two squares, every element in bS is the sum of two squares.

Chapter 33

Reading Course

A reading course for projective geometry, based in large part on these notes. Use any source of help you can find. Collaboration is recommended.

33.1 Incidence Structures

Read Section 1.1 on incidence structures.

1. Suppose $\mathcal{P}$ is a projective plane, possibly degenerate. Prove that the following are equivalent:
   (a) $\mathcal{P}$ contains a 4-arc.
   (b) The incidence graph of $\mathcal{P}$ is thick.
2. Prove that an incidence structure is a projective plane if and only if its incidence graph is bipartite with diameter three and girth six.
3. Determine the eigenvalues of the incidence graph of a (non-degenerate) projective plane.
4. Show that the set of points and lines fixed by a group of collineations of a projective plane is a projective plane, possibly degenerate. Determine the degenerate projective planes.
5. Let p be a point and ℓ a line in the projective plane $\mathcal{P}$ of order n. Let X be the subgraph of the incidence graph of $\mathcal{P}$ induced the points not on ℓ and the lines not on p. Show that X is an antipodal distance-regular cover of $K_{n.n}$. [The converse also holds.]

6. A *generalized quadrangle* is a partial linear space such that if p is a point and ℓ is a line not on q, there is a unique point on ℓ collinear with p. If thick, the incidence graph is semiregular (see e.g., G&R: AGT). Determine what the possibilities are if the incidence graph is not thick.

7. Prove that an incidence structure is a generalized quadrangle if and only if its incidence graph has diameter four and girth eight.

8. Let $\mathbb{E}$ be an extension field of $\mathbb{F}$ with degree three and assume $q = |\mathbb{F}|$. Then $\mathbb{E}$ is a 3-dimensional vector space over $\mathbb{F}$, and its 1- and 2-dimensional vector subspaces form a projective plane $\mathcal{P}$. Use the fact that the multiplicative group of a finite field is cyclic to show that there is a cyclic group of collineations of $\mathcal{P}$ of order q^2+q+1, acting regularly on the points of $\mathcal{P}$. [Hence we may assume that the incidence matrix of $\mathcal{P}$ is a circulant.]

Read Sections 4.4–5.4.

33.2 Collineations

A *collineation* of an incidence structure is an automorphism of its incidence graph that map each of the colour classes to itself. So it is a pair (P, B) of permutations such that P acts on the point and B acts on blocks and incidence is preserved. In matrix terms, if N is the incidence matrix of the structure, then (P, B) is a collineation if and only if $PNQ^T = N$.

Let $(\mathcal{P}, \mathcal{B})$ be an incidence structure with an incidence matrix N. Let ρ be a partition of $\mathcal{P}$ with characteristic matrix R. The i-th entry in the column of $R^T N$ corresponding to the block β is the number of points incident with β that lie in the i-th cell of ρ. Let ρ^* be the partition of $\mathcal{B}$, where two blocks lie in the same cell if and only if the corresponding columns of $R^T N$ are equal. We say ρ^* is the partition *induced* by ρ.

33.2.1 Theorem. *Let $(\mathcal{P}, \mathcal{B})$ be an incidence structure, let ρ be a partition of its points and let ρ^* be the induced partition of its blocks. If the rows of the incidence matrix of $(\mathcal{P}, \mathcal{B})$ are linearly independent, then $|\rho| \leq |\rho^*|$.* □

33.2.2 Corollary. *Let $\mathcal{I}$ be an incidence structure and assume that the rows of its incidence matrix are linearly independent. If Γ is a group of*

collineations of $\mathcal{I}$, the number of orbits of Γ on blocks is at least as large as the number of orbits on points. □

1. Let γ be a non-identity collineation of a projective plane of order n that fixes all points on some line ℓ. Show that either:

   (a) γ fixes exactly one point not on ℓ, and the order of γ divides $n-1$, or

   (b) γ does not fix any point off ℓ, and the order of γ divides n.

   A collineation as in (a) is known as a *homology*, in (b) we have an *elation*.

2. Characterizw the structures that arise as the set of fixed points and fixed lines of a group of collineations of a generalized quadrangle.

Index