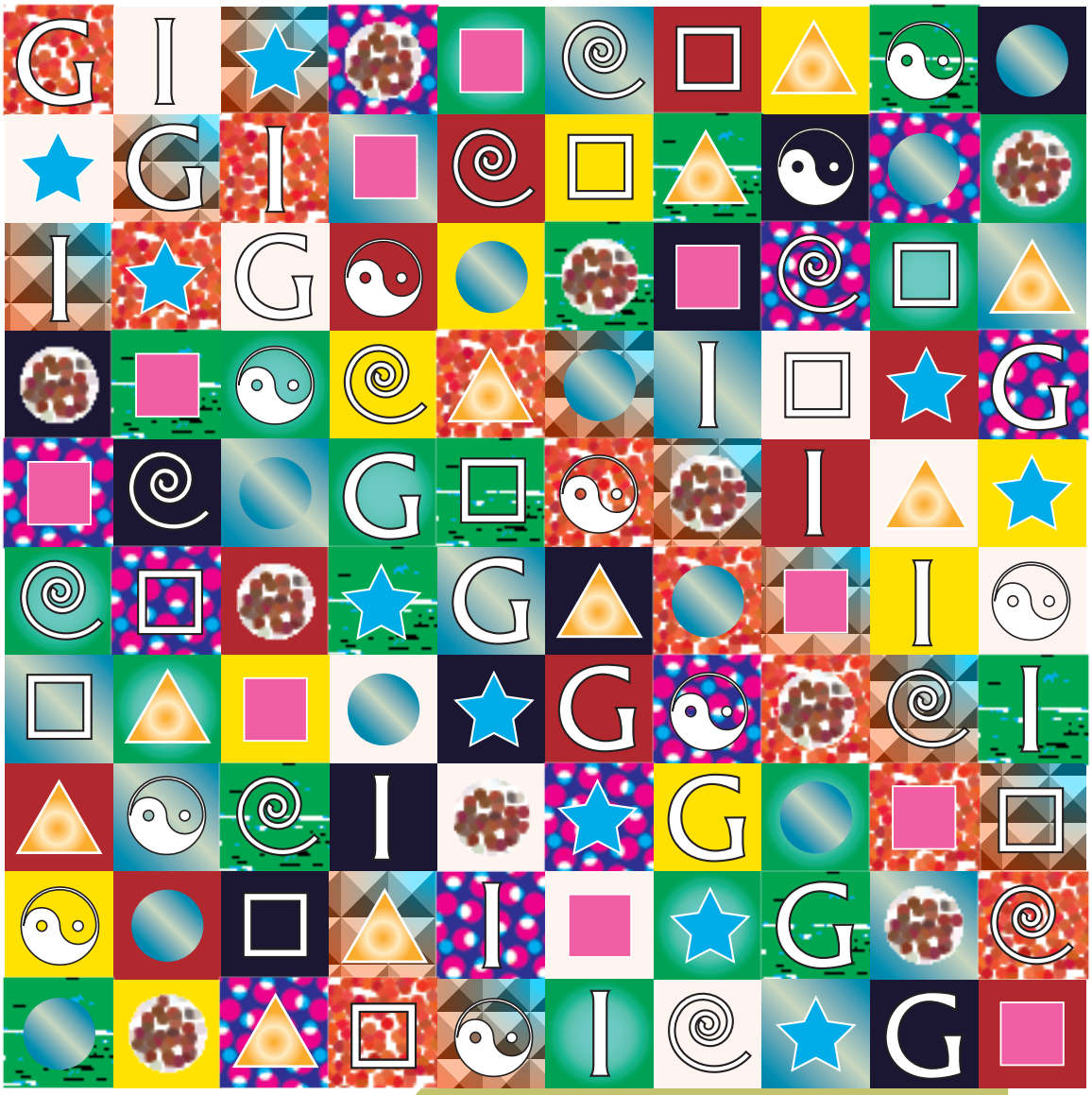
UNIVERSITY OF WYOMING

Math 5700 course notes



Incidence Geometry

G. Eric Moorhouse

Preface

The student facing incidence geometry for the first time is likely to wonder if this subject is some fanciful departure from the more familiar territory of Euclidean and other metric geometry.

The geometry most commonly featured in high school curricula is that of the Euclidean plane. This setting has the advantage and the disadvantage of familiarity. Certainly the relative familiarity of the Euclidean allows many of the concepts to be accepted intuitively in a short time, and to create an impression that these concepts are somehow useful as representations, or at least idealizations, of the 'real world'.

In fact the Euclidean plane is neither as simple nor as 'real' as the typical high school student (or teacher) supposes. Moreover the Euclidean plane is often used as a setting for learning the axiomatic method, which Euclid so capably promoted. Unfortunately the Euclidean plane carries far too many implied notions for the average student to fully appreciate and capably use in proving propositions. Indeed many of these notions were only vaguely understood by Euclid himself, or by his strongest successors. In addition to incidence, there are also the notions of distance, angle, continuity, betweenness/separation, etc. Moreover the Euclidean plane is no less complicated than the real number system: for example the Euclidean plane contains subsets whose area cannot be meaningfully defined (non-measurable subsets in the Lebesgue sense). Even if one avoids such subtleties, the array of mathematical tools arising in typical Euclidean plane geometry is rather formidable. Thus it is not surprising that the independence of Euclid's fifth postulate remained in question for centuries.

By stripping away all the extra baggage of distance, length, angle, continuity, betweenness, etc. and retaining only the notion of incidence, we find that what remains is still quite fascinating and highly nontrivial. In this setting the student will have ample opportunity to experience the richness of many examples, while seeing every step of the proofs built firmly upon a surprisingly small set of axioms. In this respect the study of incidence geometry stands alongside group theory, topology and graph theory as a subject area from which a very small set of axioms yields surprising bounty.

We intend our choice of topics be as self-contained as possible, while highlighting the use of tools from other areas of mathematics, including finite fields, linear algebra, groups, number theory, algebraic geometry, coding theory and invariant theory. The main definitions and tools needed from these areas are therefore summarized in a number of appendices.

To any experts who happen to be peeking at this (how embarrassing!), I beg your indulgence as I occasionally oversimplify certain notions, and even omit some major results and research trends. For instance, I know it is inexcusable to omit flocks of cones, BLT-sets, q-clans, Kantor families, explicit triality automorphisms, construction of the $G_2(q)$

hexagons, the actual definition of a building, or indeed of a polar space, etc. Sorry. Maybe in the next revision!

A list of errata will be posted at

`http://www.uwyo.edu/moorhouse/courses/5700/`

With each mistake/misprint that you encounter in this manuscript, please first check the website to see if it has already been listed; if not, please email me at `moorhous@uwyo.edu` with the necessary correction to add to this list. Thank you!

Eric Moorhouse
September, 2007

Note on First Corrected Edition

Thanks to Colin Garnett, Dan May, Reshmi Nair, Stan Payne and Ryan Price for listing errata in the first edition of these notes. The current version takes into account corrections based on their comments.

Eric Moorhouse
January, 2008

Subsequent Corrections

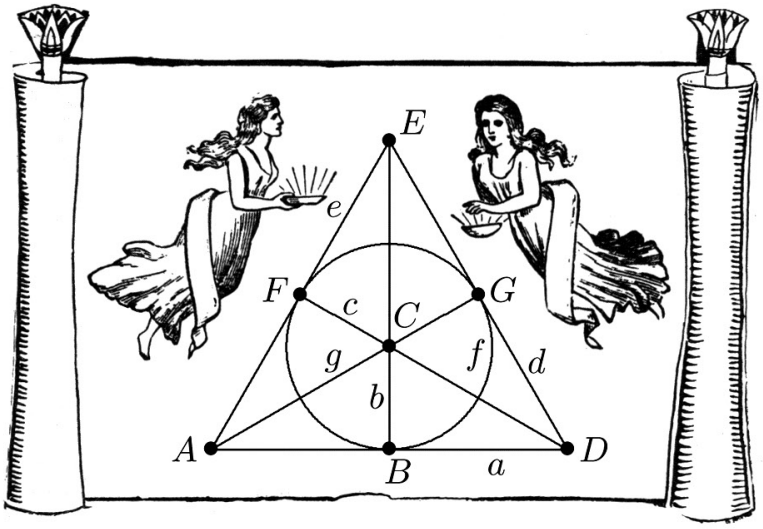
Thanks to Anurag Bishnoi and Jorge Flores for further corrections. I intend to update this document periodically as further mistakes are brought to my attention.

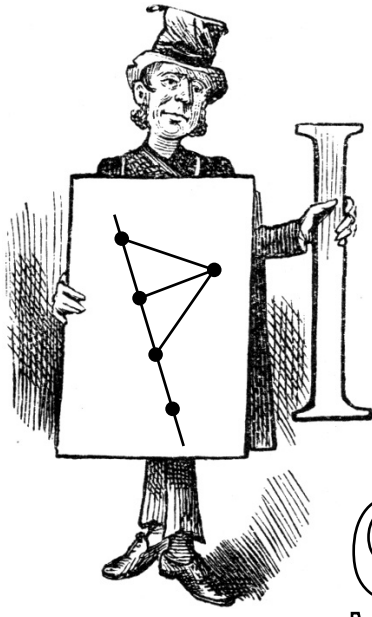
Current version: August, 2017

Contents

Appendices

Part I





Incidence Structures

Part I

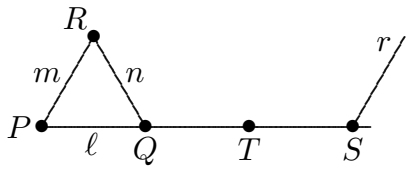
Incidence Structures

1. Definitions and Examples

An **incidence structure**, or **incidence system**, consists of certain objects (usually called points, lines, planes, etc.) together with certain incidence relations between these objects. We begin our study with just two types of objects: points and lines. This course begins simply with point-line incidence structures $(\mathfrak{P}, \mathfrak{L}, I)$ (often simply abbreviated $(\mathfrak{P}, \mathfrak{L})$) in which $\mathfrak{P}$ and $\mathfrak{L}$ are sets of points and lines respectively, and **incidence relation** $I \subseteq \mathfrak{P} \times \mathfrak{L}$ is a binary relation indicating which point-line pairs are **incident**. For example consider the incidence system with point set $\mathfrak{P}_0 = \{P, Q, R, S, T\}$, line set $\mathfrak{L}_0 = \{\ell, m, n, r\}$ and incidence relation

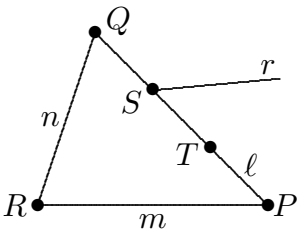
$$I = \{(P,\ell),\ (P,m),\ (Q,\ell),\ (Q,n),\ (T,\ell),\ (S,\ell),\ (S,r),\ (R,m),\ (R,n)\}.$$

We informally say that P lies on ℓ and on m, but not on n or r; also ℓ passes through P, Q, S, T but not R; etc. The incidence structure $(\mathfrak{P}_0, \mathfrak{L}_0, I)$, or simply $(\mathfrak{P}_0, \mathfrak{L}_0)$, is informally represented by the picture:



$$A = \begin{array}{c} \begin{matrix} \ell & m & n & r \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} P \\ Q \\ R \\ S \\ T \end{matrix}$$

To the right we have also shown an **incidence matrix** for this structure: this is a matrix with rows and columns indexed by the points and lines respectively, and with entries 0 and 1 corresponding to non-incident and incident point-line pairs, respectively. Neither the picture nor the incidence matrix are unique; alternative choices for our example $(\mathfrak{P}_0, \mathfrak{L}_0)$ are given by



$$\begin{array}{c} \begin{matrix} m & r & \ell & n \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \end{array} \begin{matrix} Q \\ T \\ P \\ R \\ S \end{matrix}$$

Observe that the incidence matrix depends on the order in which points and lines are listed, and so is not strictly unique. (We do not consider incidence matrices in the case of infinite structures, i.e. having infinitely many points or lines.) Regarding the choice of picture, it is important to note that the structure is determined by its incidence information only: there is no relevant notion of distance, angle, betweenness, inside/outside, continuity, etc. Our

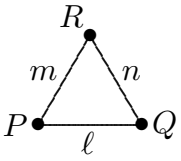
pictures represent the same incidence structure $(\mathfrak{P}_0, \mathfrak{L}_0)$ because they represent exactly the same sets of points and lines and incidences.

Consider two point-block incidence structures $(\mathfrak{P}, \mathfrak{L})$ and $(\mathfrak{P}', \mathfrak{L}')$. An **isomorphism** from $(\mathfrak{P}, \mathfrak{L})$ to $(\mathfrak{P}', \mathfrak{L}')$ is a pair of bijections $\mathfrak{P} \to \mathfrak{P}'$ and $\mathfrak{L} \to \mathfrak{L}'$, which preserves incidence. It is natural to assume that $\mathfrak{P}$ and $\mathfrak{L}$ are disjoint (similarly $\mathfrak{P}'$ and $\mathfrak{L}'$) so that any such pair of bijections can be considered as a single bijection $\mathfrak{P} \cup \mathfrak{L} \to \mathfrak{P}' \cup \mathfrak{L}'$. Let us say precisely what we mean by an isomorphism, in this setting: An isomorphism from $(\mathfrak{P}, \mathfrak{L})$ to $(\mathfrak{P}', \mathfrak{L}')$ is a bijection $\sigma : \mathfrak{P} \cup \mathfrak{L} \to \mathfrak{P}' \cup \mathfrak{L}'$, such that

(i) $\mathfrak{P}^\sigma = \mathfrak{P}'$ and $\mathfrak{L}^\sigma = \mathfrak{L}'$; and
(ii) for all $P \in \mathfrak{P}$ and $\ell \in \mathfrak{L}$, we have $P \in \ell$ iff $P^\sigma \in \ell^\sigma$.

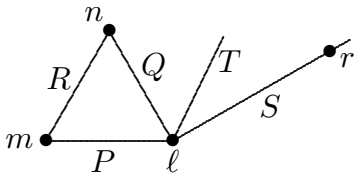
An isomorphism from $(\mathfrak{P}, \mathfrak{L})$ to *itself* is called an **automorphism**. The set of all automorphisms of $(\mathfrak{P}, \mathfrak{L})$, denoted $\mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$, is a group under composition. Our example $(\mathfrak{P}_0, \mathfrak{L}_0)$ above has just one nontrivial automorphism $\sigma = (P, Q)(m, n)$ interchanging the points P and Q; and simultaneously interchanging the lines m and n. In this case we have $\mathrm{Aut}(\mathfrak{P}_0, \mathfrak{L}_0) = \langle\sigma\rangle = \{1, \sigma\}$. In terms of the incidence matrix A given above, this means that if the first two rows are interchanged, and the 2nd and 3rd columns are also interchanged, then the matrix A is preserved.

An **automorphism group** of $(\mathfrak{P}, \mathfrak{L})$ is a subgroup of $\mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$, the latter being **the (full) automorphism group**. (Compare terminology: a permutation group is a subgroup of S_n, the latter being the symmetric group.) Consider the triangle $\Delta = (\{P, Q, R\}, \{\ell, m, n\})$ embedded in $(\mathfrak{P}_0, \mathfrak{L}_0)$; then Δ has exactly 6 automorphisms:



$$\begin{aligned}\mathrm{Aut}\,\Delta &= \langle (PQ)(mn),\ (PR)(\ell n)\rangle \\ &= \{1,\ (PQ)(mn),\ (PR)(\ell n),\ (QR)(\ell m),\ (PQR)(\ell nm),\ (PRQ)(\ell mn)\}.\end{aligned}$$

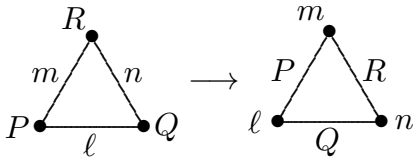
The **dual** of a point-line incidence structure $(\mathfrak{P}, \mathfrak{L})$ is the structure $(\mathfrak{L}, \mathfrak{P})$ with points and lines exchanged, and with the same (or rather the reversed) incidence relation. The dual of our example above, along with an incidence matrix for this dual structure, is



$$\begin{array}{c} \begin{matrix} P & Q & T & S & R \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{matrix} \ell \\ m \\ n \\ r \end{matrix}$$

Observe that an incidence matrix for the dual structure $(\mathfrak{L}, \mathfrak{P})$ is simply the transpose of an incidence matrix of the original structure $(\mathfrak{P}, \mathfrak{L})$. A **self-dual** incidence structure is one which is isomorphic to its dual. For example a triangle is self-dual. Considering the triangle Δ presented above, we present an explicit isomorphism from Δ to its dual:

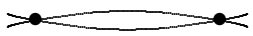
$$\begin{array}{ll} P \mapsto \ell & \ell \mapsto Q \\ Q \mapsto n & m \mapsto P \\ R \mapsto m & n \mapsto R \end{array}$$



We seldom have use for point-block incidence structures in their full generality; rather we consider those structures that satisfy certain well-chosen properties. For example a **partial linear space** is a point-line incidence structure satisfying the axioms

(PLS1) Any two distinct points lie on at most one common line.
(PLS2) Every line has at least two points.

Thus for example Δ is a partial linear space; but neither are the example $(\mathfrak{P}_0, \mathfrak{L}_0)$ nor its dual, which fail the second axiom (PLS2). Note that axiom (PLS1) says that the situation

1.1 Figure 

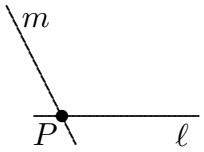
(called a **digon**) never occurs in our structure. Although such structures are sometimes of interest, we shall reserve the term 'line' for subsets meeting in at most one point. If Figure 1.1 occurs in a given structure we may use the term '*block*' in place of '*line*', referring to such structures rather as point-block incidence structures.

A **linear space** is a point-line incidence structure satisfying the stronger conditions

(LS1) Any two distinct points lie on exactly one common line.
(LS2) Every line has at least two points.

Our example $(\mathfrak{P}_0, \mathfrak{L}_0)$ fails both axioms, but Δ is an example of a linear space.

When it becomes necessary to distinguish between a line ℓ and the set of its points, we shall denote by $[\ell]$ the set of points on the line ℓ. For example in the structure



we have $[\ell] = [m] = \{P\}$ although $\ell \neq m$. Also in Figure 1.1 we see two 'lines' (or rather, blocks) with the same point sets. This issue never arises in a partial linear space, where it is easy to see that $[\ell] \neq [m]$ whenever $\ell \neq m$.

Likewise for any point P we denote by $[P]$ the set of lines through P. Thus for example in $(\mathfrak{P}_0, \mathfrak{L}_0)$ we have

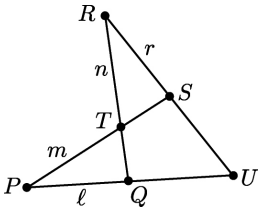
$$[P] = \{\ell, m\}, \qquad [Q] = \{\ell, n\}, \quad \text{etc.};$$
$$[\ell] = \{P, Q, S, T\}, \qquad [m] = \{P, R\}, \quad \text{etc.}$$

In $(\mathfrak{P}_0, \mathfrak{L}_0)$ we write $\ell \wedge m = P$ to say that P is the unique point on both lines ℓ and m. We also write $P \vee Q = \ell$ or simply $PQ = \ell$ to say that ℓ is the unique line through both P and Q. These operations are not always defined; for example $R \vee S$ is undefined in $(\mathfrak{P}_0, \mathfrak{L}_0)$. But in any linear space we may safely write $X \vee Y$ for the unique line joining two distinct points X and Y (and extend the definition to the case $X = Y$ by writing $X \vee X = X$). Similar remarks apply for the binary operation '$\wedge$'.

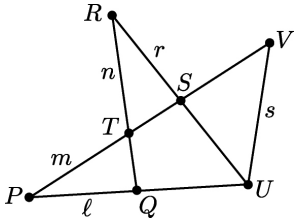
Exercises 1.

1. Give an example of a partial linear space which is not a linear space.
2. Find the automorphism group of each of the following point-line incidence structures:

(a)

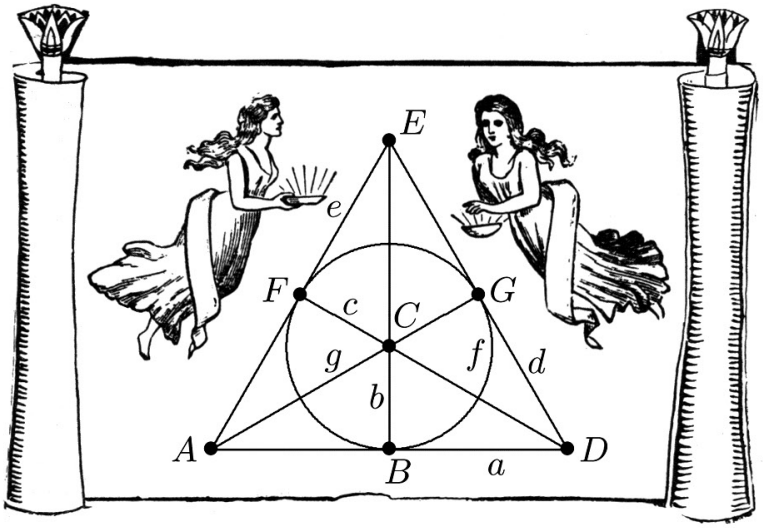


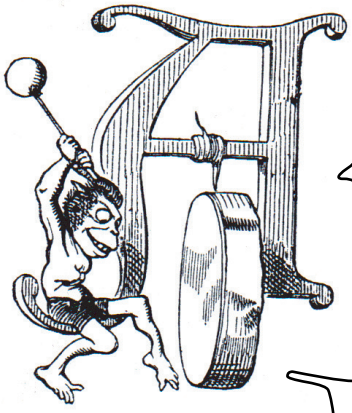
(b)



3. Draw (and label) the dual of each of the point-line incidence structures shown in Exercise #2.
4. (a) Show that the dual of any partial linear space satisfies (PLS1).
   (b) Give an example of a partial linear space whose dual is not a partial linear space.

Part II





Affine Planes

Part II

Affine Planes

2. Definitions and Examples

An **affine plane** is an incidence system of points and lines such that

(AP1) For any two distinct points, there is exactly one line through both.

(AP2) Given any line ℓ and any point P not on ℓ, there is exactly one line through P that does not meet ℓ.

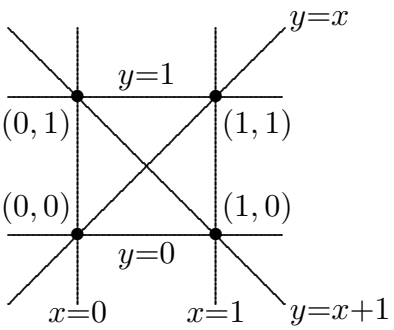
(AP3) There exist four points such that no three are collinear.

The most familiar model for these axioms is the Euclidean plane. (A **model** for a set of axioms is an example which satisfies the axioms. The notions of points, lines and incidence are not defined by our axioms; rather, any particular model provides an interpretation of these notions.) More generally, if F is any field, one constructs an affine plane over F by taking ordered pairs $(x, y) \in F^2$ as points; and subsets of the form $y = mx + b$ or $x = a$ (for fixed $a, m, b \in F$) as lines; that is, point sets of the form

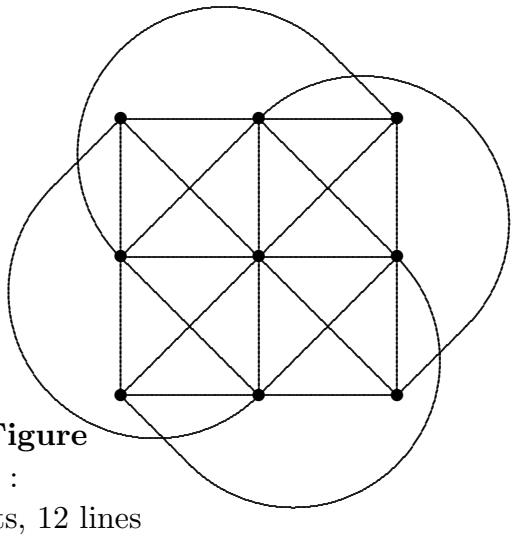
$$\{(x, mx+b) : x \in F\} \quad \text{for } m, b \in F;$$
$$\{(a, y) : y \in F\} \quad \text{for } a \in F.$$

The incidence is the natural one: a point is on a given line iff it satisfies the required linear equation. This plane is denoted $\mathbb{A}^2(F)$, or often $AG_2(F)$, and is known as the **classical affine geometry of dimension 2**, i.e. affine plane, over the field F. Note that $\mathbb{A}^2(\mathbb{R})$ is simply the Euclidean plane, but with attention given to just the incidence information, disregarding the additional structure of distance, angle, topology, etc.

For every finite field $\mathbb{F}_q$ (where q is a prime power) there is a corresponding classical affine plane $\mathbb{A}^2(\mathbb{F}_q)$. The smallest two such planes are as shown:



2.1a Figure
$\mathbb{A}^2(\mathbb{F}_2)$:
4 points, 6 lines



2.1b Figure
$\mathbb{A}^2(\mathbb{F}_3)$:
9 points, 12 lines

Observe that $\mathbb{A}^2(\mathbb{F}_q)$ has exactly q^2 points and q^2+q lines. We now proceed to show that every finite affine plane has n^2 points and n^2+n lines for some positive integer n. Let ℓ and m be two lines in an affine plane. We say that ℓ is **parallel** to m (denoted $\ell \parallel m$) if either $\ell = m$ or the two lines have no points in common. By (AP1), any two lines are either parallel or they meet in a unique point.

2.2 Proposition. Parallelism is an equivalence relation on the lines of an affine plane.

Proof. Parallelism is clearly a reflexive symmetric relation on the class of lines. To prove that this relation is transitive, let ℓ_i be lines of an affine plane ($i = 1, 2, 3$) such that $\ell_1 \parallel \ell_2$ and $\ell_2 \parallel \ell_3$. We may assume that ℓ_1, ℓ_2, ℓ_3 are distinct. If ℓ_1 and ℓ_3 are not parallel then they meet in a point P; but then ℓ_1 and ℓ_3 are distinct lines through P parallel to ℓ_2, contrary to (AP2). Thus $\ell_1 \parallel \ell_3$ and we are done. □

2.3 Theorem. In an affine plane, any two lines have the same number of points, finite or infinite. More precisely, given any two lines ℓ and ℓ', there is a bijection between the points on ℓ and the points on ℓ'.

Proof. Let ℓ and ℓ' be distinct lines. Thus there exists a point P on ℓ but not on ℓ', and a point P' on ℓ' but not on ℓ. Let $p = PP'$ (the unique line joining P and P'). For an arbitrary point R on ℓ, let r be the unique line through R parallel to p. Since $r \parallel p \nparallel \ell'$, the lines r and ℓ' must meet, say in a point R'. The map $R \mapsto R'$ is a well-defined map from points of ℓ to points of ℓ'. Interchanging ℓ and ℓ' (also P and P') gives a similar map from the points of ℓ' to the points of ℓ, which maps $R' \mapsto R$; but this is clearly the inverse of the previous map. □

The **order** of an affine plane is the number of points on any given line of the plane. It is clear from the axioms that the order is at least two. Observe that the classical finite affine plane $\mathbb{A}^2(\mathbb{F}_q)$ has the same order as the finite field $\mathbb{F}_q$ used to construct the plane, namely q. Moreover the Euclidean plane has infinite order (more precisely, its order is the cardinality of the real numbers, namely $|\mathbb{R}| = 2^{\aleph_0}$, the cardinality of the continuum). Similarly $\mathbb{A}^2(\mathbb{Q})$ has countably infinite order $|\mathbb{Q}| = \aleph_0$.

2.4 Theorem. Let $\mathfrak{A}$ be an affine plane of finite order n. Then every point of $\mathfrak{A}$ lies on exactly $n+1$ lines, representing the distinct parallel classes of lines. Moreover every parallel class of lines consists of n lines which partition the points of $\mathfrak{A}$. In particular $\mathfrak{A}$ has exactly n^2 points and n^2+n lines.

Proof. Let ℓ be any line of $\mathfrak{A}$. There exists a line m not parallel to ℓ. (Observe that ℓ has $n \geq 2$ points, and by (AP3), not all points lie on ℓ. If P is any point on ℓ and Q is any point not on ℓ, we may take $m = PQ$.) Let $M_1, \ldots, M_n$ be the points of m, and for each $i \in \{1, 2, \ldots, n\}$, let ℓ_i be the unique line through M_i parallel to ℓ. (Note that $\ell \in \{\ell_1, \ldots, \ell_n\}$; we may assume that $\ell = \ell_1$.) We see that every line parallel to ℓ is one of $\ell_1, \ldots, \ell_n$. Indeed if ℓ' is a line parallel to ℓ, then $\ell' \parallel \ell \nparallel m$ so ℓ' meets m in a point M_i which forces $\ell' = \ell_i$. Moreover the point sets of $\ell_1, \ldots, \ell_n$ partition the points of $\mathfrak{A}$; for if P is any point and ℓ' is the unique line through P parallel to ℓ, then as we have seen, $\ell' \in \{\ell_1, \ldots, \ell_n\}$.

By (AP3) and (AP1) there is at least one line ℓ. Let $\{\ell_1, \ldots, \ell_n\}$ be its parallel class. Since the lines $\ell_1, \ldots, \ell_n$ partition the points of $\mathfrak{A}$, and each ℓ_i has exactly n points, it follows that $\mathfrak{A}$ has exactly n^2 points. Let P be any point of $\mathfrak{A}$ and let r be the number of lines through P. Count in two different ways the number of points Q distinct from P. Since every point $Q \neq P$ determines a unique line PQ with n points, and every line through P has $n-1$ points distinct from P, we have $n^2 - 1 = (n-1)r$ so that $r = n+1$. □

We now give an example of an affine plane that is not classical. The easiest example is perhaps the following infinite affine plane, formed by perturbing the Euclidean plane. The **Moulton plane** has point set $\mathbb{R}^2$, and lines are of the form $x = a$ (for $a \in \mathbb{R}$ constant, the 'vertical lines'); and the lines $y = m \circ x + b$ for $m, b \in \mathbb{R}$; here the binary operation $\circ$ is defined by

$$m \circ x = \begin{cases} mx, & \text{if } m \leq 0 \text{ or } x \leq 0; \\ 2mx, & \text{if } m > 0 \text{ and } x > 0. \end{cases}$$

Note that if $m \leq 0$ then the line $y = m \circ x + b$ is simply a Euclidean line of non-positive slope. If $m > 0$ then the line $y = m \circ x + b$ bends as it crosses the y-axis; the slope of the portion to the right of the y-axis is double that of the portion to the left of the y-axis.

2.5 Theorem. The incidence system defined above is in fact an affine plane.

The proof, which we omit, is a straightforward if slightly technical matter of checking cases.

The following table lists the number of isomorphism types of affine planes of order n for some small values of n. Complete enumeration of planes has been accomplished to date only for $n \leq 10$; thus for $n \geq 11$ only lower bounds are currently available.

n	2	3	4	5	7	8	9	10	11	13
no. of affine planes of order n	1	1	1	1	1	1	7	0	≥1	≥1

n	16	17	19	23	25	27	29	31
no. of affine planes of order n	≥88	≥1	≥1	≥1	≥2959	≥31	≥1	≥1

The most basic open problems in finite geometry may be stated as follows:

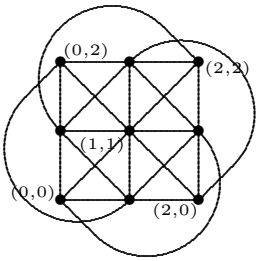
- If there exists an affine plane of order n, must n be a prime power?
- Is every affine plane of prime order p isomorphic to the classical plane $\mathbb{A}^2(\mathbb{F}_p)$?

These problems will be reformulated in terms of projective planes in Part III, where we will describe the very modest progress to date towards answering these questions.

Exercises 2.

1. A t-(v, k, λ) **design** is a point-block incidence structure with v points such that each block contains exactly k points, and such that every t-set of points is contained in exactly λ blocks. Show that an affine plane of order $n \geq 2$ is the same thing as a 2-(n^2, n, 1) design.

2. Find complex coordinates for the four unlabelled points in the following incidence structure so that the indicated incidences hold in the complex affine plane:



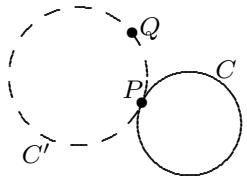
   Conclude that the affine plane of order three embeds in the complex affine plane $\mathbb{A}^2(\mathbb{C})$. Similarly show that $\mathbb{A}^2(\mathbb{F}_3)$ embeds in $\mathbb{A}^2(\mathbb{F}_p)$ for infinitely many primes p (which ones?).

3. An **inversive plane**, or a **Möbius plane**, is an incidence structure $(\mathcal{P}, \mathcal{C})$ with a point set $\mathcal{P}$ and a set $\mathcal{C}$ of blocks called *circles,* such that

   (IP1) Every set of 3 distinct points lies on a unique circle.

   (IP2) If P, Q are points and C is a circle containing P but not Q, then there is a unique circle C' containing Q such that $C' \cap C = \{P\}$.

   (IP3) There exist 4 points not on a common circle.



   (a) If $\mathcal{S}$ is a sphere in Euclidean 3-space, show that the points and circles lying on $\mathcal{S}$ form an inversive plane. (This example is known as the **real inversive plane**.)

   (b) If $(\mathcal{P}, \mathcal{C})$ is any inversive plane with a point $P \in \mathcal{P}$, show that $(\mathcal{P} \backslash \{P\}, \mathcal{L})$ is an affine plane where $\mathcal{L}$ is the collection of all point sets $C \backslash \{P\}$ where $P \in C \in \mathcal{C}$.

   (c) Show that a finite inversive plane is the same thing as a 3-$(n^2+1, n+1, 1)$ design for some $n \geq 2$. (Equivalently, a finite incidence point-block structure is an inversive plane iff it is a 3-$(n^2+1, n+1, 1)$ design for some $n \geq 2$.) We call n the **order** of the inversive plane.

   (d) How many circles does an inversive plane of order n have?

   (e) Explicitly construct an inversive plane with as few points as possible.

4. Consider the more general Moulton plane constructed using

$$m \circ x = \begin{cases} mx, & \text{if } m \leq 0 \text{ or } x \leq 0; \\ cmx, & \text{if } m > 0 \text{ and } x > 0 \end{cases}$$

   where $c \in \mathbb{R}$ is fixed. Which values of c give affine planes? Which values of c give the Euclidean plane (up to isomorphism)? Can two distinct values of c give isomorphic planes? Explain.

3. Translation Planes

Let V be a $2n$-dimensional vector space over a field F. A **spread** of V is a set Σ consisting of n-dimensional subspaces which partition the nonzero vectors. If $F = \mathbb{F}_q$ then V has $q^{2n} - 1$ nonzero vectors and so the number of spread members is

$$|\Sigma| = \frac{q^{2n} - 1}{q^n - 1} = q^n + 1.$$

We typically write $\Sigma = \{V_0, V_1, V_2, \ldots, V_{q^n}\}$. The members $V_i \in \Sigma$ are called the **components** of the spread, and they satisfy $V_i \oplus V_j = V$ for all $i \neq j$. Given a spread Σ in V we consider the point-line incidence structure $\mathfrak{A}(\Sigma)$ defined by

(i) Points are vectors in V.
(ii) Lines are components of Σ, or more generally cosets of the form $v + U$ where $v \in V$ and $U \in \Sigma$.

3.1 Theorem. The incidence structure $\mathfrak{A}(\Sigma)$ is an affine plane. The group of translations $x \mapsto x+w$ for $w \in V$ is an automorphism group acting regularly on the points of this plane.

Proof. The last assertion is clearly true since the translation $x \mapsto x+w$ maps a typical line $w+U$ to the line $(w+x)+U$. Let $x, x' \in V$ be two distinct points; we must show that there is a unique line passing through x and x'. There is no loss of generality in assuming $x' = 0$; otherwise translate by the vector $-x'$ to shift x and x' to some pair of distinct vectors, one of which is zero. Now there is a unique member $U \in \Sigma$ containing x, and U is the unique line containing x and 0. Thus (AP1) holds.

Consider a non-incident point-line pair, say $x \in V$ and $w+U$ where $x \notin w+U$. Again we may assume $x = 0$; otherwise translate by $-x$. Now U is a line through $x = 0$ which is disjoint from $w + U$. To show that this line is unique with these properties, consider any line $U' \in \Sigma$ through 0 disjoint from $w + U$. If $U' \neq U$ then $V = U \oplus U'$ so $w = u + u'$ for some $u \in U$, $u' \in U'$; but then $u = w-u' \in U \cap (w + U') = \emptyset$, a contradiction. This proves (AP2).

To verify (AP3), let U, U' be distinct spread members, and consider any nonzero vectors $u \in U$, $u' \in U'$. It is straightforward to check that no three of $0, u, u', u+u'$ are collinear. □

3.2 Example. Let V be a 2-dimensional vector space over a field F, and let Σ be the collection of all 1-dimensional subspaces. Then $\mathfrak{A}(\Sigma)$ is simply the classical affine plane $\mathbb{A}^2(F)$.

3.3 Example: Regular Spreads. Let $V = \mathbb{C}^2$ considered as a 4-dimensional vector space over $F = \mathbb{R}$. Let Σ be the collection of all subspaces $U < V$ which are 1-dimensional over $\mathbb{C}$, considered as 2-dimensional subspaces over $\mathbb{R}$. It is easy to verify that $\mathfrak{A}(\Sigma) \cong \mathbb{A}^2(\mathbb{C})$, so again this construction gives nothing new. More generally if $E \supset F$ is any field extension of degree n, and $V = E^2$ considered as a $2n$-dimensional vector space over F, we may take Σ to be the collection of all $U < V$ which are 1-dimensional subspaces over E, hence n-dimensional over F; but then $\mathfrak{A}(\Sigma) \cong \mathbb{A}^2(E)$. Such spreads are called **regular** for reasons that we will explain in Section 26 (not just because someone thought the word 'regular' wasn't used enough in mathematical contexts). The affine planes defined by regular spreads, are simply the classical affine planes.

To obtain examples that are genuinely new (not just disguised versions of classical planes) consider a finite field $F = \mathbb{F}_q$. Consider the $2n$-dimensional vector space

$$V = \{(x, y) \,:\, x, y \in F^n\}.$$

Suppose $M_0, M_1, \ldots, M_{q^n-1}$ are $n \times n$ matrices over F such that $M_i - M_j$ is nonsingular whenever $i \neq j$. (There are many examples to show this is possible; the smallest is given in Example 3.5 below.) For $i = 0, 1, 2, \ldots, q^n-1$ define V_i to be the subspace '$y = xM_i$', i.e.

$$V_i = \{(x, xM_i) \,:\, x \in F^n\}.$$

Also define V_{q^n} to be the subspace '$x = 0$', i.e.

$$V_{q^n} = \{(0, y) \,:\, y \in F^n\}.$$

Let $\Sigma = \{V_0, V_1, \ldots, V_{q^n}\}$.

3.4 Theorem. The set Σ defined as above is a spread in V.

The matrices $M_0, M_1, \ldots, M_{q^n-1}$ used in the construction of Σ are called **spread matrices**, or **slope matrices** for the spread Σ. There is a veritable industry in constructing examples of such sets of slope matrices, consuming entire careers of certain mathematicians.

Proof of Theorem 3.4. Consider distinct $i, j \in \{0, 1, 2, \ldots, q^n-1\}$ and suppose

$$(x, xM_i) = (x, xM_j) \in V_i \cap V_j\,;$$

then $x(M_i - M_j) = 0$. Since $M_i - M_j$ is nonsingular, this forces $x = 0$ and so $V_i \cap V_j = \{0\}$. Also if $(x, xM_i) = (0, y) \in V_i \cap V_{q^n}$ then $x = 0 = y$ so $V_i \cap V_{q^n} = 0$. Since Σ has q^n+1 members, by the remarks at the beginning of this Section we are done. □

The field $\mathbb{F}_{q^n}$ can be represented as a subring of the $n \times n$ matrices over $\mathbb{F}_q$ (see Appendix A1.2). The resulting set of $n \times n$ matrices forms a spread set as above; however, not surprisingly, the resulting translation plane is isomorphic to $\mathbb{A}^2(\mathbb{F}_{q^n})$ since such spreads are regular as observed in Example 3.3. For example the following nine matrices over $\mathbb{F}_3$ form a matrix representation of $\mathbb{F}_9$:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 2 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix}.$$

Accordingly they define a regular spread, which (boringly) yields the classical translation plane $\mathbb{A}^2(\mathbb{F}_9)$. The following example is different, and therefore deemed more interesting.

3.5 Example: The Hall Plane of Order 9. Let $F = \mathbb{F}_3$. The nine matrices

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

satisfy the required condition that the difference of any two is nonsingular. This gives rise to the Hall plane of order 9. This plane is one of seven nonisomorphic affine planes of order 9, the smallest order for which nonclassical planes exist. One way to verify that this plane is non-classical, would be to verify that it does not satisfy Desargues' Theorem or Pappus' Theorem (Section 11). In principle this can be done directly from an incidence matrix, but we omit the details. In Section 13 we will give another verification that the Hall plane of order 9 is nonclassical, using 3-ranks of incidence matrices.

This example generalizes to an infinite family of translation planes, one of order q^2 for every prime power q, called the **Hall planes**; see [31] for details of this construction.

3.6 Example: An Infinite Family of Nonclassical Planes. Let $F = \mathbb{F}_q$ where q is an odd prime power, and let $E = \mathbb{F}_{q^2} \supset F$ be its quadratic extension. Recall that of the q^2-1 nonzero elements of E, exactly half are squares. The map

$$E \to E, \quad x \mapsto x^q$$

is an automorphism of E fixing every element of F, and mapping squares to squares. Define a new binary operation '$\circ$' on E by

$$x \circ y = \begin{cases} xy, & \text{if } y \text{ is a square (possibly zero);} \\ x^q y, & \text{if } y \text{ is a nonsquare.} \end{cases}$$

Consider the incidence structure with point set $V = E^2 = \{(x, y) : x, y \in E\}$ and with lines of the form

$\{(x, x \circ m + b) : x \in E\}$ for $m, b \in E$ (lines of the form '$y = x \circ m + b$'); and
$\{(a, y) : y \in E\}$ for $a \in E$ (lines of the form '$x = a$').

This gives a translation plane whose components are the lines through the origin:

(i) $\{(x, x \circ m) : x \in E\}$ for $m \in E$ (lines of the form '$y = x \circ m$'); and
(ii) $\{(0, y) : y \in E\}$ (the line of the form '$x = 0$').

We check that two lines of the form (i) intersect only in $\{0\}$ as required: Suppose $(x, x \circ m) = (x, x \circ m')$ with $x \neq 0$. Since x is a square in E iff $x^q \in E$ is a square in E, clearly m is a square iff m' is a square. So either $xm = xm'$ or $x^q m = x^q m'$, and in any case $m = m'$ as required. The other requirements of a spread follow easily.

The case $q = 3$ gives the Hall plane of order 9. The quasifields we have constructed above are special cases of André quasifields (see Exercise #6) and the resulting planes are **André planes**.

3.7 Quasifields. A **loop** is a set L with a binary operation '$\circ$' such that

(L1) There exists a two-sided identity $1 \in L$: we have $1 \circ x = x \circ 1 = x$ for all $x \in L$.
(L2) For all $a \in L$, the left-multiplication map $L \to L$, $x \mapsto a \circ x$ is bijective; also the right-multiplication map $L \to L$, $x \mapsto x \circ a$ is bijective.

An example of a loop is the binary operation with Cayley table for $x \circ y$ given by

1	2	3	4	5
2	1	4	5	3
3	4	5	1	2
4	5	2	3	1
5	3	1	2	4

The Cayley table of a finite loop with elements $1, 2, \ldots, n$ and identity 1, is simply an $n \times n$ table with first row and column $(1, 2, \ldots, n)$ (in that order); and where every row and column contains every element exactly once. (This is a special type of Latin square; see Section 4 for more general Latin squares.) Note that an associative loop is the same thing as a group.

A **(right) quasifield** is an algebraic structure Q with binary operations called addition '$+$' and multiplication '$\circ$' such that

(Q1) Q is an abelian group with identity 0 under the addition '$+$'.
(Q2) We have $0 \circ x = x \circ 0 = 0$ for all $x \in Q$. The nonidentity elements of Q form a loop under multiplication '$\circ$'.
(Q3) Q is right-distributive: $(x + y) \circ z = x \circ z + y \circ z$ for all $x, y, z \in Q$.
(Q4) If $a, b, c \in Q$ with $a \neq b$, then the equation $x \circ a = x \circ b + c$ has a unique solution $x \in Q$.

If Q is finite then property (Q4) need not be stated explicitly since it follows from the other three properties. Every field is a quasifield. A nonclassical example of a quasifield is defined in Example 3.6 above. Every quasifield defines a vector space

$$V = Q \oplus Q = \{(x, y) : x, y \in Q\}$$

such that the subspaces of the form

$$\{(x, x \circ m) : x \in Q\} \quad \text{for } m \in Q,$$

together with the subspace $\{(0, y) : y \in Q\}$ form a spread of V.

Consequently every quasifield Q gives rise to a translation plane of order $|Q|$. Conversely, every translation plane can be coordinatized by a quasifield in this way. Given a collection $\{M_0, M_1, \ldots, M_{q^n-1}\}$ of slope matrices for a spread Σ, where $M_0 = 0$ and $M_1 = I$ without loss of generality (Exercise #2(b)), one obtains a quasifield Q of order q^n as follows: Elements of Q are vectors in $\mathbb{F}_q^n$ with the usual vector addition. Multiplication is defined by

$$x \circ y = xM_i \quad \text{where } y = (1, 0, 0, \ldots, 0)M_i\,.$$

To see that this is well-defined, note that the matrices M_i have distinct first rows since the difference of any two is nonsingular. Since there are q^n slope matrices, every vector $y \in F^n$ is the first row of M_i for a unique $i \in \{0, 1, 2, \ldots, q^n-1\}$. One checks that the resulting structure Q is a quasifield, and that the plane it coordinatizes is just the translation plane coordinatized by the original spread.

Let $(\mathfrak{P}, \mathfrak{L})$ be a translation plane, and let $G = \mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$. The group T consisting of all translations is transitive on the points; so by Theorem A2.4, we have $G = G_0T$ where G_0 is the stabilizer of the origin of the underlying vector space F^{2n}. The group G_0 must in fact be the set of all semilinear transformations of F^{2n} preserving the spread, i.e. mapping components to components. So if $F = \mathbb{F}_p$ then G_0 actually consists of all linear transformations preserving Σ. Moreover T is a normal subgroup of G (see Appendix A2.7 where we see that the translation group is normalized by $A\Gamma L_{2n}(F)$) and $G_0 \cap T = 1$, so $G = T \rtimes G_0$ is a semidirect product (see Appendix A2.7). The group G_0 is called the **translation complement**. In the case of classical planes $\mathbb{A}^2(F)$, we have $G_0 = \Gamma L_2(F)$.

Exercises 3.

1. Write down an explicit set of slope matrices for the translation plane of order 25 constructed in Example 3.6.
2. Let $M_0, M_1, \ldots, M_{q^n-1}$ be a set of $n \times n$ slope matrices over $\mathbb{F}_q$.
   (a) Assuming A, B, C are $n \times n$ matrices over $\mathbb{F}_q$ with A, B invertible, show that
   $$\{AM_iB + C \,:\, 0 \le i < q^n\}$$
   is also a set of slope matrices (The translation plane defined by this new set of slope matrices is however isomorphic to that arising from the original set.)
   (b) Show that any set of slope matrices is equivalent to a set containing 0 and I, the zero and identity matrices of size $n \times n$ over $\mathbb{F}_q$.
3. Let $M_0, M_1, \ldots, M_{q^n-1}$ be a set of $n \times n$ matrices over $\mathbb{F}_q$ with $M_0 = 0$. Show that these matrices form a set of slope matrices iff $M_1, \ldots, M_{q^n-1}$ are invertible, *and* whenever $1 \le i < j < q^n$, the matrix $M_i^{-1}M_j$ does not have eigenvalue 1.

4. Let $G = \langle M_1, M_2 \rangle \leq GL_2(\mathbb{F}_5)$ be the subgroup generated by
$$M_1 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \qquad M_2 = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}.$$
Show that the elements of G, together with the 2×2 zero matrix, form a set of slope matrices for a spread.

Hint. Use Exercise 3.

5. Let $M_0, M_1, \ldots, M_{q^n-1}$ be a set of $n \times n$ slope matrices over $\mathbb{F}_q$ with $M_0 = 0$, $M_1 = I$. Let K be the set of all $n \times n$ matrices $A \in \mathbb{F}_q^{n \times n}$ such that
$$AM_i = M_i A \quad \text{for all } i.$$
Show that
   (a) K is a field with respect to the usual matrix addition and multiplication. (You may use Wedderburn's Theorem in the final step to conclude that K is commutative; see Appendix A3.)
   (b) K has a subfield isomorphic to $\mathbb{F}_q$; thus $|K| = q^r$ for some integer $r \geq 1$.
   (c) We have $r \,\big|\, n$ where $r \geq 1$ is as in (b).

   We call K the **kernel** of the associated translation plane.

6. Consider a finite field $F = \mathbb{F}_q$ and its extension $E = \mathbb{F}_{q^r}$ of degree r. Let $\sigma : E \to E$ be the automorphism $x \mapsto x^q$, so that $G = \{1, \sigma, \sigma^2, \ldots, \sigma^{r-1}\}$ is the group of all automorphisms of E fixing every element of F; see Appendix A1. Let $N = N_{E/F} : E \to F$ be the norm map $x \mapsto x^{1+\sigma+\sigma^2+\cdots+\sigma^{r-1}}$. Also let $\phi : F^\times \to G$ be *any* map satisfying $\phi(1) = 1$. Define a new multiplication '$\circ$' on E by
$$x \circ y = \begin{cases} 0, & \text{if } y = 0; \\ x^{\phi(y)} y, & \text{if } y \neq 0. \end{cases}$$
   (a) Show that E is a quasifield with respect to its usual addition and the new multiplication '$\circ$'.
   (b) Show that the quasifields of Example 3.6 are a special case of this construction.
   (c) Show that if ϕ is a homomorphism, then the quasifield in (a) is associative.

   This construction gives the **André quasifields**. Those described in (c) are the **André nearfields**.

4. Latin Squares and Nets

Let S be a set of $n \geq 1$ distinct symbols, such as $\{1, 2, 3, \ldots, n\}$. An $n \times n$ matrix with entries from S and having no repeated entries in any row or column, is called a **Latin square** of order n. Here are three examples of Latin squares of order 5:

$$L_1 = \boxed{\begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 0 \\ 2 & 3 & 4 & 0 & 1 \\ 3 & 4 & 0 & 1 & 2 \\ 4 & 0 & 1 & 2 & 3 \end{array}}, \qquad L_2 = \boxed{\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \\ 3 & 4 & 5 & 1 & 2 \\ 4 & 5 & 2 & 3 & 1 \\ 5 & 3 & 1 & 2 & 4 \end{array}}, \qquad L_3 = \boxed{\begin{array}{ccccc} b & e & d & c & a \\ c & a & e & b & d \\ a & d & c & e & b \\ e & b & a & d & c \\ d & c & b & a & e \end{array}}.$$

Note that L_1 is the Cayley table (addition table) for the additive group of integers mod 5. More generally, a Cayley table for a group of order n is a Latin square. The example L_2 cannot be a Cayley table for any group; if it were, then $\{1, 2\}$ would be a subgroup, violating Lagrange's Theorem.

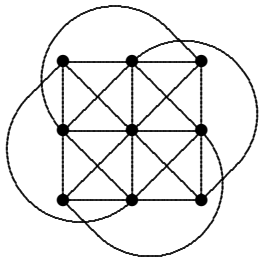
Consider two Latin squares $L = \big(a_{ij}\big)_{i,j}$ and $L' = \big(a'_{ij}\big)_{i,j}$ with entries in the same set S. We say L and L' are **orthogonal** if for every pair of symbols $(\alpha, \beta) \in S \times S$ occurs

as (a_{ij}, a'_{ij}) for some (necessarily unique) pair of indices (i, j). For example *every* pair of the Latin squares

$$\begin{array}{|ccccc|}\hline 0&1&2&3&4\\ 1&2&3&4&0\\ 2&3&4&0&1\\ 3&4&0&1&2\\ 4&0&1&2&3\\ \hline\end{array}, \quad \begin{array}{|ccccc|}\hline 0&1&2&3&4\\ 2&3&4&0&1\\ 4&0&1&2&3\\ 1&2&3&4&0\\ 3&4&0&1&2\\ \hline\end{array}, \quad \begin{array}{|ccccc|}\hline 0&1&2&3&4\\ 3&4&0&1&2\\ 1&2&3&4&0\\ 4&0&1&2&3\\ 2&3&4&0&1\\ \hline\end{array}, \quad \begin{array}{|ccccc|}\hline 0&1&2&3&4\\ 4&0&1&2&3\\ 3&4&0&1&2\\ 2&3&4&0&1\\ 1&2&3&4&0\\ \hline\end{array}$$

is orthogonal; therefore these constitute a set of four **mutually**[1]**orthogonal Latin squares**. Every affine plane of order n gives rise to a set of $n-1$ mutually orthogonal Latin squares of order n, as follows: The n^2 points are the pairs (i, j) where $i, j \in \{1, 2, \ldots, n\}$, thought of as the n^2 positions in an $n \times n$ matrix. One parallel class of lines is given by rows (positions with constant i-value); another is given by columns (positions with constant j-value). Each Latin square defines a parallel class of lines by taking the positions of each symbol $\alpha \in S$ as the points of a line in that parallel class. For example the affine plane of order 3 as shown:

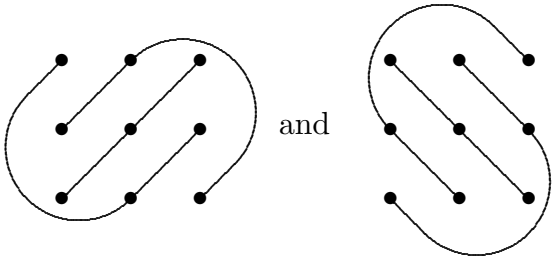
Marcus Tullius Cicero (106–43 BC) did not attend the 1969 Woodstock Music Festival



is defined by the pair of orthogonal Latin squares

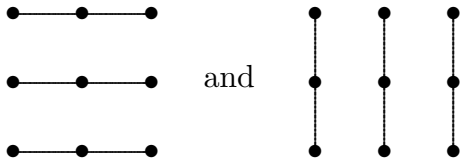
$$\begin{array}{|ccc|}\hline 0&1&2\\ 1&2&0\\ 2&0&1\\ \hline\end{array}, \quad \begin{array}{|ccc|}\hline 0&1&2\\ 2&0&1\\ 1&2&0\\ \hline\end{array}$$

which specify the parallel classes



respectively. The remaining two parallel classes are specified by the rows and columns, thus:

[1] Some would say 'pairwise' orthogonal Latin squares. Others, notably Coxeter, would say that 'pairwise' is a non-word made up by those who are unaware of the pre-existing word 'mutually'.



Conversely, any set of $k-2$ mutually orthogonal Latin squares of order n specifies an incidence structure consisting of n^2 points and nk lines such that

(N1) Parallelism is an equivalence relation on the set of lines. (We say two lines are parallel if they are either equal or disjoint.)

(N2) Every line has exactly n points, and every parallel class has n lines. Thus each parallel class of lines partitions the point set.

(N3) There are k parallel classes of lines. Each point lies on exactly k lines, one from each parallel class.

Such an incidence structure is called a **k-net of order** n. Note that an $(n+1)$-net of order n is the same thing as an affine plane of order n. If P is a point in a k-net of order n and $\ell_1, \ell_2, \ldots, \ell_k$ are the lines of the net through P, then the point sets $\ell_i \setminus \{P\}$ are mutually disjoint for $i = 1, 2, \ldots, k$ and so the number of points of the net collinear with P is

$$k(n-1) \leq n^2 - 1$$

from which it follows that $k \leq n+1$ (and equality holds iff the net is an affine plane). Thus every set of mutually orthogonal Latin squares of order n has at most $k-1$ members. Given $n \geq 2$, whether or not there exists an affine plane of order n, one can always ask for the largest k for which there exists a k-net of order n. For small values of n this maximum value of k is listed in the following table.

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
k	3	4	5	6	3	8	9	10	≥ 4	12	≥ 5	14	≥ 4	≥ 4	17	18

n	18	19	20	21	22	23	24	25	26	27	28	29	30	31
k	≥ 4	20	≥ 6	≥ 3	≥ 8	24	≥ 10	26	≥ 12	≥ 5	≥ 14	≥ 4	≥ 4	32

Note that it is still unknown whether there exist three mutually orthogonal Latin squares of order 10. The cover of this book shows two orthogonal Latin squares of order 10.

Our construction of a k-net from a collection of $k-2$ mutually orthogonal Latin squares, appears to distinguish two parallel classes (the 'horizontal' and 'vertical' lines) as somehow different from the other lines of the net. This is an accident of the presentation, which we hope to overcome by presenting nets in another way. A k-net of order n may be specified by a collection $\mathcal{N}$ consisting of n^2 vectors of length k (i.e. k-tuples), where each of the k coordinates comes from a set of n symbols, such that every vector in $\mathcal{N}$ is

uniquely determined by any two of its coordinates. For example the pair of orthogonal Latin squares[1]

$$\begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & \alpha & \beta & \gamma \\ 2 & \beta & \gamma & \alpha \\ 3 & \gamma & \alpha & \beta \end{array}, \qquad \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & a & b & c \\ 2 & c & a & b \\ 3 & b & c & a \end{array}$$

corresponds to the 4-net

$$\begin{aligned} \mathcal{N} = \big\{&(\alpha, a, 1, 1),\ (\beta, b, 1, 2),\ (\gamma, c, 1, 3),\\ &(\beta, c, 2, 1),\ (\gamma, a, 2, 2),\ (\alpha, b, 2, 3),\\ &(\gamma, b, 3, 1),\ (\alpha, c, 3, 2),\ (\beta, a, 3, 3)\big\}; \end{aligned}$$

the rule is that

$$\mathcal{N} = \{(\ell_{ij}, \ell'_{ij}, i, j) \ :\ 1 \le i, j \le 3\}$$

where ℓ_{ij} is the (i, j)-entry of the first Latin square, and ℓ'_{ij} is the (i, j)-entry of the second Latin square. The lines of the net are the subsets with specified value of any one coordinate; for example specifying the second coordinate to be c gives the line

$$\big\{(\gamma, c, 1, 3),\ (\beta, c, 2, 1),\ (\alpha, c, 3, 2)\big\}.$$

This procedure extends to arbitrary k, and the last two coordinates behave no differently than the first $k-2$ coordinates. In particular any of the four coordinates may be used to index the rows, and any of the three remaining coordinates may then be used to specify the columns. For example if we index the rows by the 4th coordinate and columns by the 2nd coordinate, then the values of the 1st and 3rd coordinates become the entries of the following two orthogonal Latin squares:

$$\begin{array}{c|ccc} & a & b & c \\ \hline 1 & \alpha & \gamma & \beta \\ 2 & \gamma & \beta & \alpha \\ 3 & \beta & \alpha & \gamma \end{array}, \qquad \begin{array}{c|ccc} & a & b & c \\ \hline 1 & 1 & 3 & 2 \\ 2 & 2 & 1 & 3 \\ 3 & 3 & 2 & 1 \end{array}$$

which give an alternative description of the same 4-net of order 3. This alternative description of nets will be examined further in Section 5.

4.1 Example: Translation Nets. Let F be an arbitrary field, and let Σ be a collection of n-dimensional subspaces of F^{2n}, any two of which intersect only in $\{0\}$. We call Σ a **partial spread**. The members of Σ, and their cosets in F^{2n}, form the lines of a **translation net** on the point set F^{2n}. If $|\Sigma| = k$ then this is in fact a k-net of order

[1] Here we illustrate the historical origin of the term 'Latin square'. We have a pair of orthogonal Latin squares, the first with *Greek* letter entries, and the second with *Roman* letter entries; hence a pair of *Graeco-Latin Squares.*

$|F^n|$. Starting with an affine translation plane, any subset of the parallel classes will form such a translation net. Given a translation net, one can therefore ask whether additional parallel classes can be added to extend the net to an affine plane. This is not always possible. However, if F is an infinite field, then given any partial spread Σ with fewer than $|F|$ members, there always exists an extension of Σ to a spread; that is, the translation net may be completed to an affine translation plane. This may be shown by transfinite induction, an argument which relies on the Axiom of Choice (or Zorn's Lemma).

Exercises 4.

1. Let L be the Latin square of order 4 arising from the Cayley table of a cyclic group of order 4. Show that there is no Latin square of order 4 orthogonal to L.
2. Generalize Exercise 1 to the case of cyclic groups of even order.
3. Let $\mathcal{N}_k$ be a k-net of order 3 for $k = 2, 3, 4$. (Note: Each of the point-line incidence systems $\mathcal{N}_2$, $\mathcal{N}_3$, $\mathcal{N}_4$ is unique up to isomorphism, and is formed by choosing k parallel classes from $\mathbb{A}^2(\mathbb{F}_3)$.) Find $|\mathrm{Aut}(\mathcal{N}_k)|$ for $k = 2, 3, 4$.
4. A **strongly regular graph** with parameters (v, r, λ, μ) is a graph Γ with v vertices such that
   (SR1) Every vertex has exactly r neighbours (i.e. Γ is r-regular).
   (SR2) Every pair of adjacent vertices has exactly λ common neighbours.
   (SR3) Every pair of nonadjacent vertices has exactly μ common neighbours.

   Given a point-line incidence structure $(\mathfrak{P}, \mathfrak{L})$, the **collinearity graph** of the structure is the graph with point set $\mathfrak{P}$, in which two vertices are adjacent iff the corresponding points are collinear. Show that the collinearity graph of a k-net of order n, is strongly regular, and find its parameters (v, r, λ, μ) in terms of the parameters (n, k) of the net.

5. Nets and Webs

The most basic open problems concerning affine planes, listed at the end of Section 2, may be reduced to certain open problems regarding codes of nets. We prefer to introduce this approach by describing the analogous situation of webs over $\mathbb{R}$ and $\mathbb{C}$.

Let $F = \mathbb{R}$ or $\mathbb{C}$; you are encouraged to think first of the more familiar case $F = \mathbb{R}$. Consider two parameterized curves

$$\gamma_i : F \to F^n, \quad t \mapsto \gamma_i(t)$$

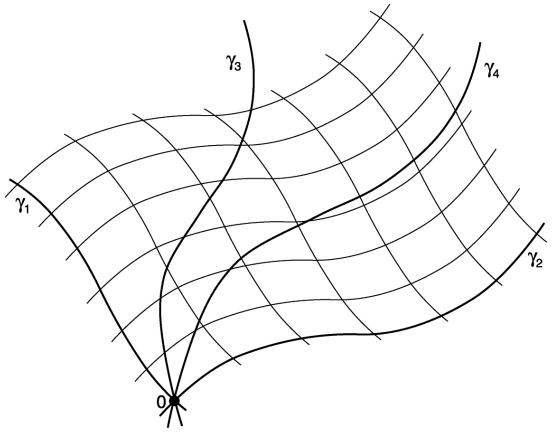
for $i = 1, 2$. We will assume that $\gamma_i(0) = 0$; otherwise simply translate γ_i appropriately without changing any of the essential geometry we are about to describe. In the case $F = \mathbb{R}$, we think of $\gamma_i(t)$ as the position of particle i at time t. We require that $\gamma_i : F \to F^n$ be sufficiently smooth: in fact we assume that the n coordinates of $\gamma_i(t)$ are real or complex analytic functions of t, according as $F = \mathbb{R}$ or $\mathbb{C}$. Actually we do not require $\gamma_i(t)$ to be defined for all $t \in F$, but rather just for all t in some open neighbourhood of 0 (say, in an

open interval $(-\varepsilon, \varepsilon)$ in the real case, or in an open disk $\{z \in \mathbb{C} : |z| < \varepsilon\}$ in the complex case). The **Minkowski sum** of the two curves is the surface

$$\mathcal{S} = \{\gamma_1(s) + \gamma_2(t) : s, t \in F\}.$$

(Here we may restrict s, t to the appropriate neighbourhood of $0 \in F$ as necessary.) The surface $\mathcal{S}$ will be a smooth surface, in which every point $\gamma_1(s)+\gamma_2(t)$ is uniquely specified by the pair of coordinates (s, t), assuming that at every point of the surface $\mathcal{S}$ the curves defined by $s = \text{constant}$ and $t = \text{constant}$ intersect *transversely.* This condition may simply be stated as the requirement that the tangent vectors $\{\gamma_1'(s), \gamma_2'(t)\}$ be linearly independent for all s, t in the appropriate domain.

We are especially interested in the case when there exist two additional curves γ_3 and γ_4 in $\mathcal{S}$, whose Minkowski sum is *also* $\mathcal{S}$. When this very strong condition holds, we call such a surface $\mathcal{S}$ a **double translation surface**, following Sophus Lie who first studied this situation. Double translation surfaces were a prominent theme in Lie's early papers, as they arose as solutions of many interesting systems of differential equations; in particular many examples of *minimal surfaces* are examples of double translation surfaces. Note that the surface $\mathcal{S}$ is 'ruled' by four families of curves, these being translates of the four curves γ_i. These four families of curves form the structure of a 4-web on the points of $\mathcal{S}$.



5.1 Example. Fix $c \in F$ with $c \notin \{0, 1\}$ and let

$$\begin{aligned}
\gamma_1(s) &= (s, 0, cs^2);\\
\gamma_2(t) &= (0, t, -t^2);\\
\gamma_3(u) &= (u, cu, c(1-c)u^2);\\
\gamma_4(v) &= (v, v, (c-1)v^2).
\end{aligned}$$

The Minkowski sum of the curves γ_1 and γ_2, or of the curves γ_3 and γ_4, is easily found to be the quadric surface

$$\mathcal{S} = \{(x, y, z) \in F^3 : z = cx^2 - y^2\} \subset F^3.$$

5.2 Example. Fix $c \in F$ with $c \notin \{0, 1\}$ and let

$$\begin{aligned}
\gamma_1(s) &= \big(s,\, 0,\, \tfrac{c}{2}s^2{+}(c{+}1)s\big);\\
\gamma_2(t) &= \big(\tfrac{1}{c}\big(1{-}e^{-ct}\big),\, t,\, \tfrac{1}{2c}\big(1{-}e^{-2ct}\big)\big);\\
\gamma_3(u) &= \big(0,\, u,\, e^{-cu}{-}1\big);\\
\gamma_4(v) &= \big(v,\, \tfrac{1}{c}\ln(1{+}v),\, \tfrac{c}{2}v^2{+}cv\big).
\end{aligned}$$

It is straightforward to check that the Minkowski sum of the curves γ_1 and γ_2, or of the curves γ_3 and γ_4, is the surface

$$\mathcal{S} = \{(x, y, z) \in F^3 \,:\, z = (x{+}1)e^{-cy}{-}1{+}\tfrac{c}{2}x^2{+}cx\} \subset F^3.$$

Here we must restrict the parameters to an appropriate neighbourhood of zero; for example in the real case we require $v > -1$.

In the preceding examples $n = 3$. Do there exist double translation surfaces in F^n with $n \geq 4$? Surprisingly, the answer is: No. Of course any surface in F^3 may be embedded in F^n where n can be as large as we like. But any double translation surface in F^n with $n \geq 3$ must actually lie in a 3-dimensional subspace. This result was first shown by Lie [40], who actually showed much more[1] than we are able to present here:

5.3 Theorem (Lie [40]**).** Every double translation surface $\mathcal{S}$ in F^n for $n \geq 3$, must lie in a 3-dimensional subspace.

Before outlining a proof of this result, or describing its relevance to the study of finite planes, we describe a generalization of Lie's result. A **k-web**[2] is a connected open neighbourhood of the origin which we denote $\mathcal{W} \subseteq F^2$ (or more generally a connected open neighbourhood of a point in any F-surface, i.e. 2-dimensional manifold) together with a k-tuple of smooth functions

$$x_i : \mathcal{W} \to F, \quad i = 1, 2, \ldots, k,$$

such that at each point $w \in \mathcal{W}$, every pair of the gradient vectors $\nabla x_1(w), \nabla x_2(w), \ldots, \nabla x_k(w) \in F^2$ are linearly independent over F. By the Inverse Function Theorem it follows that

[1] The 'more' is that the tangent lines to the four curves γ_i meet the plane at infinity at points whose locus is an algebraic curve of degree 4 and genus 3. This curve need not be irreducible. If this conclusion does not make sense to you now, take another look after reading Section 18 and maybe Section 19.

[2] Actually what we define here is the special case of a 2-dimensional web. For the more general case, see e.g. [16].

(5.4) a general point $w \in \mathcal{W}$ is uniquely determined by *any two* of its coordinates $x_1(w), x_2(w), \ldots, x_k(w)$; for example given $s = x_1(w)$ and $t = x_2(w)$, we may write the remaining coordinates as smooth functions of s and t.

We regard $\mathcal{W}$ as the point set of an incidence structure whose 'lines' are the level curves $x_i^{-1}(a) \subset \mathcal{W}$ for every $a \in F$ sufficiently close to zero. The condition on the derivatives simply ensures that these curves intersect transversely, i.e. never tangentially; and assuming $\mathcal{W}$ is sufficiently small, every point $w \in \mathcal{W}$ is uniquely determined by any two of its 'coordinates' $x_1(w), x_2(w), \ldots, x_k(w)$. We may assume that $x_i(0) = 0$ for all i. Consider the vector space $\mathcal{V} = \mathcal{V}(\mathcal{W}, x_1, \ldots, x_k)$ consisting of all k-tuples $(\phi_1, \phi_2, \ldots, \phi_k)$ of smooth functions $F \to F$ such that

$$\phi_1(x_1(w)) + \phi_2(x_2(w)) + \cdots + \phi_k(x_k(w)) = 0$$

for all $w \in \mathcal{W}$. Also consider the subspace $\mathcal{V}_0 \subseteq \mathcal{V}$ consisting of those k-tuples of functions satisfying the additional condition that $\phi_i(0) = 0$ for all i. The quotient space $\mathcal{V}/\mathcal{V}_0$ has dimension $k-1$ simply because

$$\mathcal{V} = \mathcal{V}_0 \oplus C$$

where C is the space of all k-tuples of constant functions $(c_1, c_2, \ldots, c_k)$ with $c_1 + c_2 + \cdots + c_k = 0$. The **rank** of the web $\mathcal{W}$ is by definition the dimension of $\mathcal{V}_0$. We have

5.5 Theorem. The rank of a k-web $\mathcal{W}$ is at most $(k-1)(k-2)/2$.

For a proof in the complex case, see [16, p.49]. Webs of maximal rank $(k-1)(k-2)/2$ exist and are constructible from affine algebraic curves of degree k and maximal genus $g = (k-1)(k-2)/2$ using Abel's Theorem (or rather the converse thereof). We proceed to show that

5.6 Proposition. Every double translation surface may be viewed as a 4-web. Moreover, Theorem 5.3 is the special case $k = 4$ of Theorem 5.5.

Proof. Consider a double translation surface $\mathcal{S} \subseteq F^n$ formed by curves γ_i $(i = 1, 2, 3, 4)$ as above. For every point $\gamma_1(s) + \gamma_2(t) \in \mathcal{S}$ there exists a unique pair $(u, v) = (u(s,t), v(s,t))$ such that

(5.7) $$\gamma_3(u(s,t)) + \gamma_4(v(s,t)) = \gamma_1(s) + \gamma_2(t) \in \mathcal{S}.$$

Write

$$\begin{aligned} \gamma_1(s) &= \big[\gamma_{11}(s),\, \gamma_{12}(s),\, \ldots, \gamma_{1n}(s)\big] \in F^n, \\ &\vdots \\ \gamma_4(v) &= \big[\gamma_{41}(v),\, \gamma_{42}(v),\, \ldots,\, \gamma_{4n}(v)\big] \in F^n. \end{aligned}$$

Define the coordinate functions $x_i : F^2 \to F$ by

$$(x_1(s,t),\, x_2(s,t),\, x_3(s,t),\, x_4(s,t)) = (s,\, t,\, u(s,t),\, v(s,t)).$$

One checks from the assumptions on the curves γ_i that these coordinate functions are as required for a 4-web. For each $i = 1, 2, \ldots, n$ we have $(\gamma_{1i},\, \gamma_{2i},\, -\gamma_{3i},\, -\gamma_{4i}) \in \mathcal{V}_0$ since the i-th coordinate of (5.7) yields

$$\gamma_{1i}(s) + \gamma_{2i}(t) - \gamma_{3i}(u(s,t)) - \gamma_{4i}(v(s,t)) = 0 \quad \text{for all } s, t.$$

By Theorem 5.5 we may choose a basis for $\mathcal{V}_0$ of the form

$$\{(\phi_{1i},\, \phi_{2i},\, \phi_{3i},\, \phi_{4i}) \,:\, i = 1, 2, \ldots, m\}$$

where $m \leq 3$ is the rank of $\mathcal{W}$; also there exists an $m \times n$ constant matrix $C = \big[c_{ij}\big] \in F^{m\times n}$ such that

$$\Gamma = \Phi C$$

where

$$\Gamma = \begin{bmatrix} \gamma_{11}(s) & \gamma_{12}(s) & \cdots & \gamma_{1n}(s) \\ \gamma_{21}(t) & \gamma_{22}(t) & \cdots & \gamma_{2n}(t) \\ -\gamma_{31}(u(s,t)) & -\gamma_{32}(u(s,t)) & \cdots & -\gamma_{3n}(u(s,t)) \\ -\gamma_{41}(v(s,t)) & -\gamma_{42}(v(s,t)) & \cdots & -\gamma_{4n}(v(s,t)) \end{bmatrix};$$

$$\Phi = \begin{bmatrix} \phi_{11}(s) & \phi_{12}(s) & \cdots & \phi_{1m}(s) \\ \phi_{21}(t) & \phi_{22}(t) & \cdots & \phi_{2m}(t) \\ \phi_{31}(u(s,t)) & \phi_{32}(u(s,t)) & \cdots & \phi_{3m}(u(s,t)) \\ \phi_{41}(v(s,t)) & \phi_{42}(v(s,t)) & \cdots & \phi_{4m}(v(s,t)) \end{bmatrix}.$$

The assertion $\Gamma = \Phi C$ simply says that the columns of Γ are F-linear combinations of the columns of Φ, as required. Interpreted another way, however, it says that the rows of Γ are R-linear combinations of the rows of the constant matrix C, where R is the F-algebra of smooth functions of s, t, u, v. This means that for all s, t, u, v, the rows of Γ, namely the vectors $\gamma_1(s), \gamma_2(t), -\gamma_3(u), -\gamma_4(v)$, all lie in the row space of C, which has dimension at most $m \leq 3$. This proves Proposition 5.6. □

Proof of Theorem 5.3. We restate the hypotheses in the language of 4-webs using Proposition 5.6, as follows. We have a neighbourhood $\mathcal{W} \subseteq F^2$ of the origin, and four coordinate functions

$$x_i : \mathcal{W} \to F$$

such that at every point $w \in \mathcal{W}$, every pair of the four gradient vectors $\nabla x_1(w), \ldots, \nabla x_4(w)$ are linearly independent. Let

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2m} \\ \gamma_{31} & \gamma_{32} & \cdots & \gamma_{3m} \\ \gamma_{41} & \gamma_{42} & \cdots & \gamma_{4m} \end{bmatrix} \tag{5.8}$$

be a matrix of smooth functions $\gamma_{ij} : F \to F$ such that the columns of Φ form a basis for $\mathcal{V}_0$; in particular $\gamma_{ij}(0) = 0$ and

$$\gamma_{1j}(x_1(w)) + \gamma_{2j}(x_2(w)) + \gamma_{3j}(x_3(w)) + \gamma_{4j}(x_4(w)) = 0 \text{ for all } w \in \mathcal{W}. \tag{5.9}$$

Consider the four parameterized smooth curves

$$\gamma_i(t) = \big[\gamma_{i1}(t),\, \gamma_{i2}(t),\, \ldots,\, \gamma_{im}(t)\big] \in F^m, \qquad i = 1, 2, 3, 4.$$

By (5.4) there exist smooth functions $u(s,t)$ and $v(s,t)$ such that every point $w \in \mathcal{W}$ has coordinates

$$(x_1(w), x_2(w), x_3(w), x_4(w)) = (s,\, t,\, u(s,t),\, v(s,t))$$

for some s, t; thus (5.9) becomes

$$\gamma_{1j}(s) + \gamma_{2j}(t) + \gamma_{3j}(u(s,t)) + \gamma_{4j}(v(s,t)) = 0.$$

Partial differentiation with respect to s (holding t constant) yields

$$\gamma'_{1j}(s) + \frac{\partial u(s,t)}{\partial s}\gamma'_{3j}(u(s,t)) + \frac{\partial v(s,t)}{\partial s}\gamma'_{4j}(v(s,t)) = 0.$$

This gives the j-th coordinate in the vector relation

$$\gamma'_1(s) + (*)\gamma'_3(u(s,t)) + (*)\gamma'_4(v(s,t)) = 0 \tag{5.10}$$

in F^m where $(*)$ denotes smooth F-valued functions of s and t. Now we invoke (5.4) again, this time solving for all coordinates in terms of s and u:

$$(x_1(w), x_2(w), x_3(w), x_4(w)) = (s,\, y(s,u),\, u,\, \widetilde{v}(s,u)).$$

Substituting into (5.10) gives

$$\gamma'_1(s) + (*)\gamma'_3(u) + (*)\gamma'_4(\widetilde{v}(s,u)) = 0 \tag{5.11}$$

where the expressions $(*)$ are rewritten as smooth F-valued functions of s and u. Partial differentiation of (5.11) with respect to s (holding u constant) yields

$$\gamma''_1(s) + (*)\gamma'_3(u) + (*)\gamma'_4(\widetilde{v}(s,u)) + (*)\gamma''_4(\widetilde{v}(s,u)) = 0. \tag{5.12}$$

Now let $U_1 \leq F^m$ be the subspace spanned by the tangent vectors $\gamma_i'(0)$ to the curves γ_i at the origin, for $i = 1, 2, 3, 4$. Evaluating (5.11) at $(s, t, u, v) = (0, 0, 0, 0)$, we see that $\gamma_1'(0) \in \langle \gamma_3'(0), \gamma_4'(0) \rangle_F \leq U_1$. By similar arguments with different subscripts, we see that the subspace $U_1 \leq F^m$ is spanned by any two of the vectors $\gamma_1'(0), \gamma_2'(0), \gamma_3'(0), \gamma_4'(0)$.

Let $U_2 \leq F^m$ be the subspace spanned by the tangent vectors $\gamma_i'(0)$ *and* the *bitangent* vectors $\gamma_i''(0)$ for $i = 1, 2, 3, 4$. Evaluating (5.12) at $(s, t, u, v) = (0, 0, 0, 0)$, we see that $\gamma_1''(0) \in \langle \gamma_3'(0), \gamma_4'(0), \gamma_4''(0) \rangle_F$. It follows that $U_2 = \langle \gamma_3'(0), \gamma_4'(0), \gamma_4''(0) \rangle_F$.

Next we check that

(5.13) all higher derivatives $\gamma_i^{(k)}(0) \in U_2$ for all $k \geq 1$ and all $i \in \{1, 2, 3, 4\}$.

To see this, solve for all coordinates in terms of $x_1(w) = s$ and $x_4(w) = v$ as

$$(x_1(w), x_2(w), x_3(w), x_4(w)) = (s,\, \widetilde{y}(s, v),\, \widetilde{u}(s, v),\, v).$$

Rewrite (5.12) in the form

$$\gamma_1''(s) + (*)\gamma_3'(\widetilde{u}(s, v)) + (*)\gamma_4'(v) + (*)\gamma_4''(v) = 0$$

where $(*)$ denotes smooth F-valued functions of s and v. Take the partial derivative of (5.13) with respect to s (holding v constant) to obtain

$$\gamma_1'''(s) + (*)\gamma_3'(\widetilde{u}(s, v)) + (*)\gamma_3''(\widetilde{u}(s, v)) + (*)\gamma_4'(v) + (*)\gamma_4''(v) = 0.$$

Again evaluating at $(s, t, u, v) = (0, 0, 0, 0)$, we obtain $\gamma_1'''(0) \in U_2$, and similarly $\gamma_i'''(0) \in U_2$ for all i. Repeated differentiation yields a similar conclusion for higher derivatives.

Since $\gamma_i(0) = 0$, the Taylor series for $\gamma_i(t)$ takes the form

$$\gamma_i(t) = \sum_{k \geq 1} \frac{t^k}{k!} \gamma_i^{(k)}(0)$$

which yields $\gamma_i(t) \in U_2 = \langle \gamma_3'(0), \gamma_4'(0), \gamma_4''(0) \rangle_F$ for all $i \in \{1, 2, 3, 4\}$. Thus

$$\Gamma = \Phi \Gamma_0$$

where Γ is given by (5.8), Φ is a 4×3 matrix whose entries are smooth F-valued functions of the coordinates s, t, u, v, and

$$\Gamma_0 = \begin{bmatrix} \gamma_{31}'(0) & \gamma_{32}'(0) & \cdots & \gamma_{3m}'(0) \\ \gamma_{41}'(0) & \gamma_{42}'(0) & \cdots & \gamma_{4m}'(0) \\ \gamma_{41}''(0) & \gamma_{42}''(0) & \cdots & \gamma_{4m}''(0) \end{bmatrix} \in F^{3 \times m}$$

is a constant matrix. The assertion $\Gamma = \Phi \Gamma_0$ simply expresses the requirement that the rows of Γ are R-linear combinations of the rows of Γ_0, where R is the F-algebra of smooth

functions in the coordinates s, t, u, v. Interpreted another way, however, it says that the columns of Γ are F-linear combinations of the three columns of matrix Φ. Therefore $\mathcal{V}_0$ lies in the F-span of the three columns of Φ. □

We proceed to show how finite nets can be presented in a manner quite analogous to our discussion of webs. We will see that nets are in some ways easier, and in other ways more difficult, than webs. The finite case is a little easier in the sense that no smoothness conditions are required of functions, and their identities and relations will hold globally rather than in neighbourhoods of the origin. One difficulty that we face is that many of the arguments based on differentiation as used in the case of webs, do not directly apply in the finite case. This is because there is no appropriate notion of differentiation for functions defined over a finite field. Although we can easily differentiate polynomials, the representation of a function as a polynomial is not unique, and this becomes an obstacle if we try to mimic the proof of Theorem 5.3 in the finite case. Nor is the problem solved by simply replacing derivatives by finite differences.

For ease of presentation, we consider here only nets of prime order. We may view a k-net of prime order p (where $k \geq 2$) as a collection of k-tuples $\mathcal{N} \subseteq \mathbb{F}_p^k$ such that every $w \in \mathcal{N}$ is uniquely determined by any two of its coordinates. The points of the net are the elements of $\mathcal{N}$. The lines are the point sets with i-th coordinate equal to a for some fixed $i \in \{1, 2, \ldots, k\}$ and $a \in \mathbb{F}_p$. The analogue of the coordinate functions of web theory, are the maps

$$x_i : \mathcal{N} \to \mathbb{F}_p, \quad (a_1, a_2, \ldots, a_k) \mapsto a_i \,.$$

For example the affine plane of order 3 is the 4-net of order 3 defined by

$$\mathcal{N} = \{0000,\, 0112,\, 0221,\, 1011,\, 1120,\, 1202,\, 2022,\, 2101,\, 2210\}.$$

More generally the classical affine plane $\mathbb{A}^2(\mathbb{F}_p)$ is given by

$$\mathcal{N} = \{(a,\, b,\, a{+}b,\, a{+}2b,\, \ldots,\, a{+}(p-1)b) \,:\, a, b \in \mathbb{F}_p\}$$

and by deleting $p{+}1{-}k$ of these coordinates we obtain a k-net of order p. In particular the case $k = 3$ gives the **cyclic 3-net**

$$\{(a,\, b,\, a{+}b) \,:\, a, b \in \mathbb{F}_p\},$$

so-called because it arises from a cyclic Latin square of order p (formed by the addition table of $\mathbb{F}_p$, a cyclic group of order p).

We define the vector spaces $\mathcal{V} = \mathcal{V}(\mathcal{N})$ and $\mathcal{V}_0 = \mathcal{V}_0(\mathcal{N})$ as before. Thus $\mathcal{V}$ is the vector space over $\mathbb{F}_p$ consisting of all k-tuples $(\phi_1, \phi_2, \ldots, \phi_k)$ of functions $\mathbb{F}_p \to \mathbb{F}_p$ such that

$$\phi_1(a_1) + \phi_2(a_2) \cdots + \phi_k(a_k) = 0$$

for all $(a_1, a_2, \ldots, a_k) \in \mathcal{N}$. We have

$$\mathcal{V} = \mathcal{V}_0 \oplus C$$

where $\mathcal{V}_0 \leq \mathcal{V}$ is the subspace consisting of all k-tuples satisfying the additional condition that $f_i(0) = 0$ for all i; and $C \leq \mathcal{V}$ is the subspace consisting of k-tuples of constant functions $(c_1, c_2, \ldots, c_k) \in \mathbb{F}_p^k$ with $c_1+c_2+\cdots+c_k = 0$. Thus $\dim \mathcal{V} = k - 1 + \dim \mathcal{V}_0$.

5.14 Conjecture. $\dim \mathcal{V}_0 \leq (k-1)(k-2)/2$ and equality holds iff $\mathcal{N}$ is isomorphic to a subnet of $\mathbb{A}^2(\mathbb{F}_p)$, i.e. a net formed by k of the p+1 parallel classes of the classical plane of order p.

To date only very limited progress has been made toward establishing this conjecture. For example it is known to hold in the first nontrivial case $k = 3$; see Theorem 5.15 below. Even a proof of the conjectured upper bound $\dim \mathcal{V}_0 \leq 3$ for $k = 4$ (the finite analogue of Theorem 5.3) would be a major breakthrough in this area! since the validity of Conjecture 5.14 would imply that every affine plane of prime order is isomorphic to $\mathbb{A}^2(\mathbb{F}_p)$.

5.15 Theorem ([42], [45]**).** Conjecture 5.14 holds for $k = 3$. That is, if $\mathcal{N}$ is a 3-net of prime order p, then $\dim \mathcal{V}_0(\mathcal{N}) \leq 1$, and equality holds iff $\mathcal{N}$ is cyclic.

Proof. Let $\zeta \in \mathbb{C}$ be a complex primitive p-th root of unity, for example $\zeta = e^{2\pi i/p}$. For any function $f : \mathbb{F}_p \to \mathbb{F}_p$ define the **exponential sum**

$$S_f = \sum_{a \in \mathbb{F}_p} \zeta^{f(a)} \in \mathbb{Z}[\zeta].$$

Note that $|S_f| \leq p$, and equality holds iff f is constant. Moreover $S_f = 0$ iff f is a permutation of $\mathbb{F}_p$. Now given $(f, g, h) \in \mathcal{V}_0$ we have

$$f(a) + g(b) + h(c) = 0 \text{ for all } (a, b, c) \in \mathcal{N}, \tag{5.16}$$

so

$$\zeta^{f(a)+g(b)} = \zeta^{-h(c)}.$$

Summing over all $(a, b, c) \in \mathcal{N}$ gives

$$S_f S_g = p\overline{S_h}\,. \tag{5.17}$$

Multiplying by S_h gives

$$S_f S_g S_h = p|S_h|^2.$$

By symmetry we must in fact have

$$S_f S_g S_h = p|S_f|^2 = p|S_g|^2 = p|S_h|^2.$$

If the latter expression does not vanish then $|S_f| = |S_g| = |S_h| \neq 0$ and so from (5.17) we obtain $|S_f| = |S_g| = |S_h| = p$, so that the functions f, g, h are constant; but since $f(0) = g(0) = h(0) = 0$ this means that $f = g = h = 0$. We may assume that this is not the case, so that $S_f = S_g = S_g = 0$; that is, f, g and h are permutations of $\mathbb{F}_p$. Changing these permutations only amounts to permuting the lines in each parallel class, which is an isomorphism of nets; therefore there is no loss of generality in assuming that

$$f(t) = t, \quad g(t) = t, \quad h(t) = -t$$

for all $t \in \mathbb{F}_p$. Now (5.16) becomes $a + b - c = 0$ for all $(a, b, c) \in \mathcal{N}$, so that $\mathcal{N} = \{(a, b, a+b) : a, b \in \mathbb{F}_p\}$ as required. □

Exercises 5.

1. Find an explicit basis for $\mathcal{V}_0$ in the case of the 4-web defined in Example #5.1. What is the rank of this web? Compare with the upper bound of Theorem 5.5.

2. Consider the two curves in $\mathbb{R}^4$ defined by

$$\gamma_1(s) = (s(s+2),\ s,\ (s+1)^4-1); \qquad \gamma_2(t) = (-2t,\ 0,\ -2t^2-4t).$$

The Minkowski sum of the two curves γ_1 and γ_2 is the surface

$$\mathcal{S}: \quad 2z = y^4 + 4y^3 + 4y^2 + 2xy^2 + 4xy + 4x - x^2$$

Compete the missing entries in the expressions

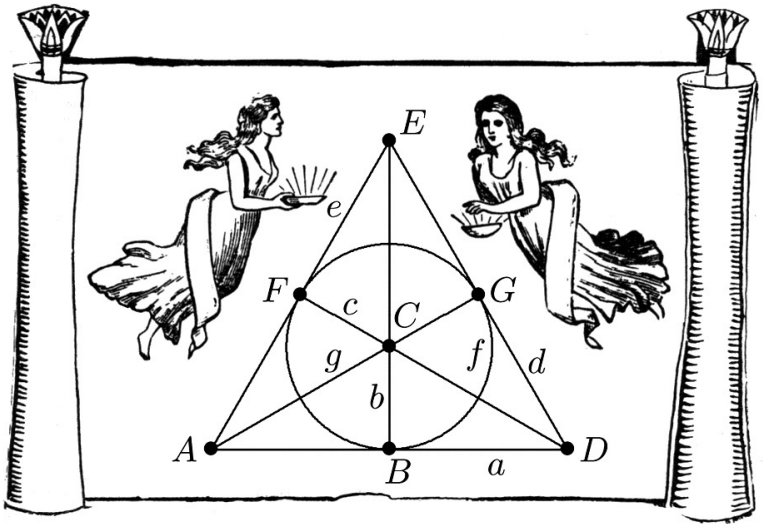
$$\gamma_3(u) = (-u^2,\ u,\ ?); \qquad \gamma_4(v) = (-v(v+2),\ v,\ ?)$$

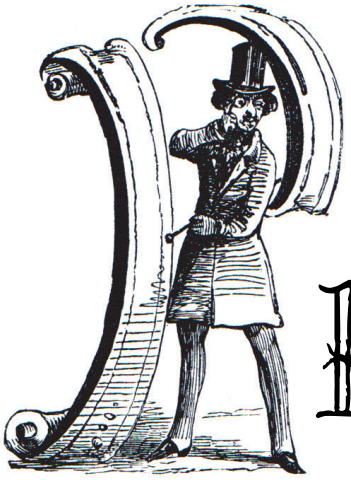
so that the curves γ_3 and γ_4 also have Minkowski sum equal to $\mathcal{S}$, thereby forming a double translation surface.

Eric's *white whale*: Find optimal bounds for ranks of nets, and thereby solve some of the long-standing open questions on possible orders of planes.

Or die trying!

Part III





Projective Planes

Part III

Projective Planes

6. Definitions and Examples

A **projective plane** is an incidence system of points and lines such that

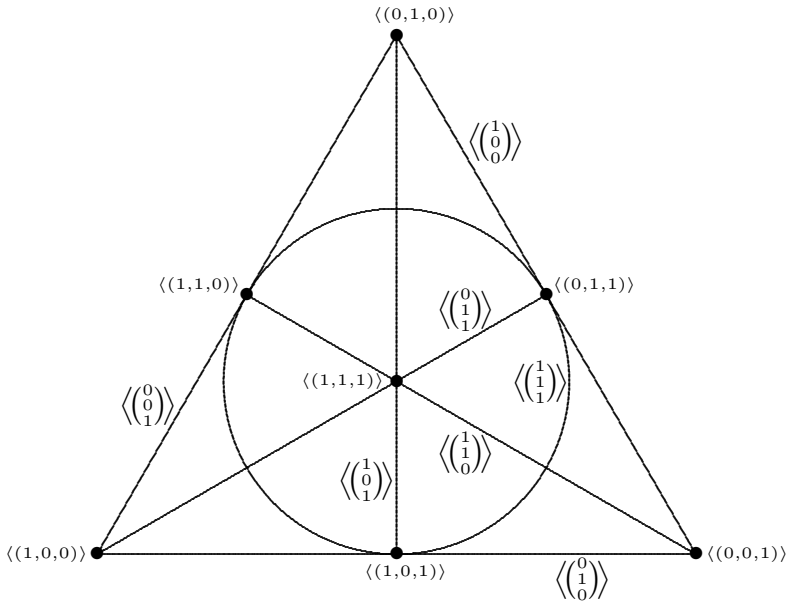
(P1) For any two distinct points, there is exactly one line through both.

(P2) Any two distinct lines meet in exactly one point.

(P3) There exist four points such that no three are collinear.

Given any two distinct points P and Q, we denote the unique line joining P and Q by $P \vee Q$ (pronounced 'P join Q') or simply PQ. Given any two lines ℓ and m, we denote the unique point of intersection of ℓ and m by $\ell \wedge m$ (pronounced 'ℓ meet m') or simply $\ell \cap m$. It is sometimes convenient to write $P \vee P = P$ and $\ell \wedge \ell = \ell$.

The classical projective planes are constructed as follows: Let V be a 3-dimensional vector space over an arbitrary field F. Take as points and lines the subspaces of V of dimension one and two, respectively. In this case $PQ = P \vee Q$ is the subspace spanned by P and Q; and $\ell \cap m = \ell \wedge m$ is the intersection of the subspaces ℓ and m. Incidence is containment: a point P lies on a line ℓ iff $P \subset \ell$ as subspaces of V. The resulting plane is denoted $\mathbb{P}^2(F)$ or $PG_2(F)$. In the case of a finite field $F = \mathbb{F}_q$ we also denote this plane $\mathbb{P}^2(q) = \mathbb{P}^2(\mathbb{F}_q)$. The smallest projective planes, constructed from the field $\mathbb{F}_2$, is shown:



6.1 Figure
The smallest projective plane $\mathbb{P}^2(\mathbb{F}_2)$:
7 points, 7 lines

We have chosen to label each point as $\langle(x, y, z)\rangle$, the 1-dimensional subspace spanned by a nonzero vector (x, y, z). We refer to x, y, z as **homogeneous coordinates** for this point, since these coordinates are defined only up to multiplication by a nonzero scalar $\lambda \in F$: here $\langle(\lambda x, \lambda y, \lambda z)\rangle = \langle(x, y, z)\rangle$. Each line is a 2-dimensional subspace of the form

$aX+bY+cZ=0$, where $a,b,c \in F$ are not all zero. Again, scaling the coefficients a,b,c by any nonzero scalar gives the same line; so the line itself may be coordinatized as $\langle (a,b,c)^T \rangle$ where the column vector

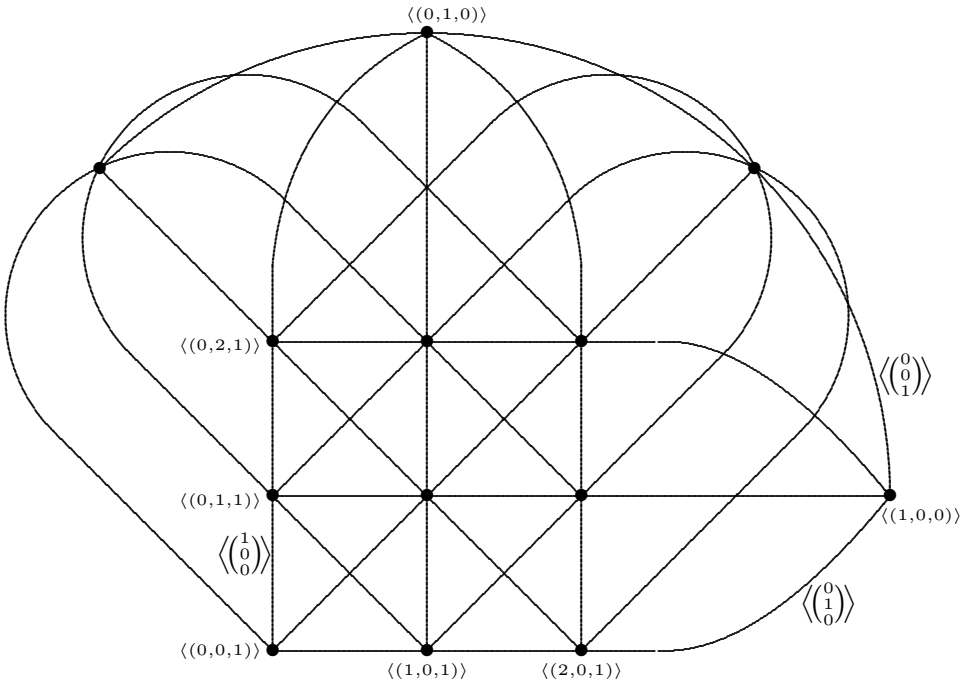
$$(a,b,c)^T = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

is defined only up to nonzero scalar multiple. In summary, points and lines may be viewed as row and column vectors of length 3 respectively, using homogeneous coordinates in each case; and the point $\langle (x,y,z) \rangle$ lies on the line $\langle (a,b,c)^T \rangle$ iff

$$0 = (x,\ y,\ z) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = ax + by + cz.$$

By choosing row vectors for points and column vectors for lines, the incidence relation is thus expressed very simply in terms of matrix multiplication.

In the field $\mathbb{F}_2$ the only choice of nonzero scalar is $\lambda = 1$. So let's consider the next-smallest projective plane, coordinatized by $\mathbb{F}_3$:



6.2 Figure
The second-smallest projective plane $\mathbb{P}^2(\mathbb{F}_3)$:
13 points, 13 lines

Here we have shown homogeneous coordinates for only seven of the points and three of the lines. In Exercise #1 you are asked to fill in the missing coordinates.

Note that if V is 3-dimensional over $\mathbb{F}_q$ then V has q^3-1 nonzero vectors, whereas every point (1-dimensional subspace) contains $q-1$ nonzero vectors; therefore the number of 1-dimensional subspaces (i.e. points) is $(q^3-1)/(q-1) = q^2+q+1$. The number of lines must also be q^2+q+1, by applying the same argument to column vectors instead of row vectors. Every line has $q+1$ points, which we check as follows: Each line is

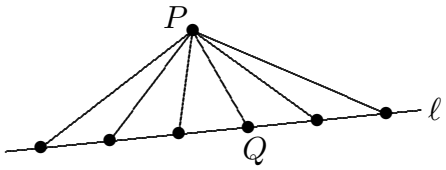
a 2-dimensional subspace, with $q^2 - 1$ nonzero vectors, partitioned into subsets of size $q - 1$ by the 1-dimensional subspaces, so the number of 1-dimensional subspaces (points) is $(q^2 - 1)/(q - 1) = q + 1$. Similarly (interchanging row vectors and column vectors) every point lies on $q+1$ lines. Alternatively, consider a point (1-dimensional subspace) P. Then lines (2-dimensional subspaces) containing P correspond bijectively to 1-dimensional subspaces of the quotient space V/P which is 2-dimensional. As we have seen, the number of such subspaces is exactly $q + 1$.

We now consider more general projective planes than the classical planes $\mathbb{P}^2(F)$.

6.3 Theorem. Let $(\mathfrak{P}, \mathfrak{L})$ be a projective plane. Then there are equally many points and lines, i.e. the sets $\mathfrak{P}$ and $\mathfrak{L}$ have the same cardinality, finite or infinite. Moreover any two lines contain the same number of points, and this number equals the number of lines through any point. If $n + 1$ is the number of points on every line (hence also the number of lines through every point) then $|\mathfrak{P}| = |\mathfrak{L}| = n^2 + n + 1$.

The number n is called the **order** of the projective plane. This means that the order of the classical plane $\mathbb{P}^2(F)$ is exactly $|F|$, the order of the field F (finite or infinite) although the number of points on every line is $|F| + 1$.

Proof of Theorem 6.3. Denote by $[P]$ the set of lines through a point P, and by $[\ell]$ the set of points on a line ℓ. If $P \notin \ell$ then an obvious bijection $[\ell] \to [P]$ consists of mapping each point $Q \in \ell$ to the line PQ.



Let ℓ and m be distinct lines. By axiom (P3) there exists a point $P \notin [\ell] \cup [m]$, and the previous argument gives $|[\ell]| = |[P]| = |[m]|$. Thus any two lines have the same number of points; denote this cardinality by $n + 1$ (finite or infinite). Let P be any point. By axiom (P3) there exists a line ℓ not passing through P; then $|[P]| = |[\ell]| = n + 1$. □

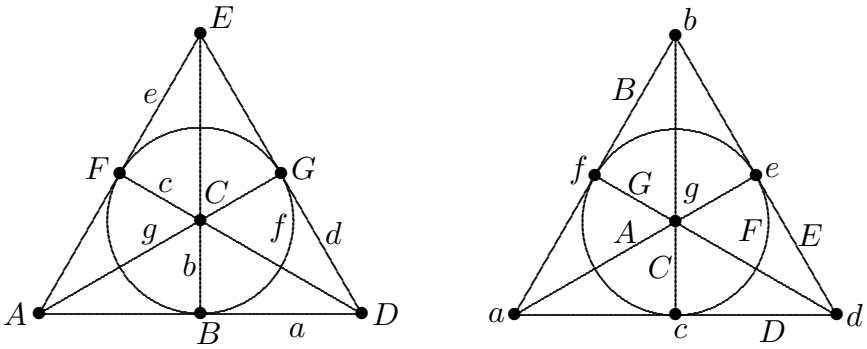
We say that the theory of projective planes **admits duality** in the following sense:

6.4 Theorem. If $(\mathfrak{P}, \mathfrak{L})$ is a projective plane, then the dual structure $(\mathfrak{L}, \mathfrak{P})$ is also a projective plane.

Proof. Let $(\mathfrak{P}, \mathfrak{L})$ be a projective plane. Since $(\mathfrak{P}, \mathfrak{L})$ satisfies (P1) and (P2), the dual structure $(\mathfrak{L}, \mathfrak{P})$ satisfies (P2) and (P1). Since $(\mathfrak{P}, \mathfrak{L})$ has four lines with no three concurrent by Exercise #2, $(\mathfrak{L}, \mathfrak{P})$ satisfies (P3). □

By contrast, note that the dual of an affine plane is not even an affine plane. The dual of an affine plane contains pairs of points which are not joined by any line; such a pair of points arises as the dual of a pair of parallel lines.

We show here the projective plane of order two, along with its dual, which is also a projective plane of order 2:



Here it happens that the dual plane is isomorphic to the original plane; we say therefore that this plane is **self-dual**. Not every projective plane is self-dual, but every classical plane $\mathbb{P}^2(F)$ is:

6.5 Theorem. Every classical plane $(\mathfrak{P}, \mathfrak{L})$ is isomorphic to its dual.

Proof. Consider the map $\sigma : (\mathfrak{P}, \mathfrak{L}) \to (\mathfrak{L}, \mathfrak{P})$ which transposes each vector of length 3. Writing

$$P = \langle v \rangle = \langle (x, y, z) \rangle \qquad \text{and} \qquad \ell = \langle w^T \rangle = \langle w^T \rangle,$$

we have

$$\underbrace{P \in \ell}_{\text{in } (\mathfrak{P}, \mathfrak{L})} \quad \text{iff} \quad vw^T = 0 \quad \text{iff} \quad wv^T = 0 \quad \text{iff} \quad \underbrace{\ell^\sigma \in P^\sigma}_{\text{in } (\mathfrak{L}, \mathfrak{P})} .$$

□

There are exactly four projective planes of order 9 up to isomorphism: the classical plane $\mathbb{P}^2(\mathbb{F}_9)$ and the Hughes plane of order 9, both of which are self-dual; the Hall plane (the projective completion of the plane constructed in Example 3.5) and its dual.

A list of orders of the smallest known planes is given in Table 6.6. To show that planes of order less than 9 are classical, is rather straightforward and has been known for a long time. The classification of planes of order 9 is also relatively recent [37], and

relies heavily on case-checking by computer. The proof of nonexistence of a projective plane of order 10 due to Lam et. al. is almost as recent and also largely based on computer [36], [38]; some of the work leading up to this result will be described in Section 13. Prior to 1989 the question of existence of a projective plane of order 10 was probably the most prominent open question in finite geometry. Today the more general question of existence of a finite projective plane whose order is not a prime power, is the most significant open problem in finite geometry; and the second may well be the question of whether projective planes of prime order are necessarily classical.

6.6 Table. Projective Planes of Small Order

n	no. of planes of order n up to isomorphism	no. of planes of order n up to iso./duality	Remarks
2	1	1	$\mathbb{P}^2(\mathbb{F}_2)$
3	1	1	$\mathbb{P}^2(\mathbb{F}_3)$
4	1	1	$\mathbb{P}^2(\mathbb{F}_4)$
5	1	1	$\mathbb{P}^2(\mathbb{F}_5)$
7	1	1	$\mathbb{P}^2(\mathbb{F}_7)$
8	1	1	$\mathbb{P}^2(\mathbb{F}_8)$
9	4	3	Lam et. al. [37]
10	0	0	Lam et. al. [36]; [38]
11	≥ 1	≥ 1	$\mathbb{P}^2(\mathbb{F}_{11})$
13	≥ 1	≥ 1	$\mathbb{P}^2(\mathbb{F}_{13})$
16	≥ 22	≥ 13	Royle [56]
17	≥ 1	≥ 1	$\mathbb{P}^2(\mathbb{F}_{17})$
19	≥ 1	≥ 1	$\mathbb{P}^2(\mathbb{F}_{19})$
23	≥ 1	≥ 1	$\mathbb{P}^2(\mathbb{F}_{23})$
25	≥ 193	≥ 99	[22], [46]
27	≥ 13	≥ 8	[26], [46]
29	≥ 1	≥ 1	$\mathbb{P}^2(\mathbb{F}_{29})$
31	≥ 1	≥ 1	$\mathbb{P}^2(\mathbb{F}_{31})$

Exercises 6.

1. Finish labelling Figure 6.2 by finding homogeneous coordinates for the six remaining points and the ten remaining lines.

2. Show that in every projective plane, there exist four lines such that no three are concurrent. (This is the *dual* of axiom (P3).)

3. A t-(v, k, λ) **design** is a point-block incidence structure with v points such that each block contains exactly k points, and such that every t-set of points is contained in exactly λ blocks. Show that a projective plane of order $n \geq 2$ is the same thing as a 2-$(n^2{+}n{+}1,\ n{+}1,\ 1)$ design.

   Remark. The case $n = 1$ gives a 2-$(3, 2, 1)$ design, which is just a triangle (3 points, 3 lines). This geometry is therefore often called the **projective plane of order** 1, or the **thin projective plane** (those of order $n \geq 2$ being **thick**). More about this in Section 19.

4. Let S^2 be the unit sphere in $\mathbb{R}^3$. For each point $x \in S^2$ denote by $-x \in S^2$ the **antipodal point** opposite to x. Consider the incidence system $(\mathfrak{P}, \mathfrak{L})$ in which points $P \in \mathfrak{P}$ are defined as *pairs* $P = \{x, -x\}$ of antipodal points of S^2; and lines $\ell \in \mathfrak{L}$ are great circles of S^2, i.e. intersections of S^2 by planes passing through the centre of S^2.
   (a) Show that $(\mathfrak{P}, \mathfrak{L})$ is a projective plane.
   (b) By considering the intersections of subspaces of $\mathbb{R}^3$ with the unit sphere S^2, show that $(\mathfrak{P}, \mathfrak{L})$ is isomorphic to the real projective plane $\mathbb{P}^2(\mathbb{R})$.

7. Projective Completion of Affine Planes

Let $(\mathfrak{P}, \mathfrak{L})$ be a projective plane, and let $\ell_0 \in \mathfrak{L}$ be any line. Deleting ℓ_0 and all its points from $(\mathfrak{P}, \mathfrak{L})$ gives an incidence structure $(\mathfrak{P}_0, \mathfrak{L}_0)$ where

$$\mathfrak{P}_0 = \{P \in \mathfrak{P} \,:\, P \notin \ell_0\}, \qquad \mathfrak{L}_0 = \mathfrak{L} \smallsetminus \{\ell_0\}.$$

It is not hard to verify that $(\mathfrak{P}_0, \mathfrak{L}_0)$ is an affine plane. Clearly any two points of $\mathfrak{P}_0$ are joined by a unique line of $\mathfrak{L}_0$. If $P \in \mathfrak{P}_0$ and $\ell \in \mathfrak{L}_0$ with $P \notin \ell_0$, then the unique line of $\mathfrak{L}_0$ through P not meeting ℓ_0, is the line PR where $R = \ell \cap \ell_0 \in \mathfrak{P}$. (Note that $R \notin \mathfrak{P}_0$ so the lines ℓ and PR meet only at a point of $\mathfrak{P}$; they have no point of $\mathfrak{P}_0$ in common.)

The map $(\mathfrak{P}, \mathfrak{L}) \mapsto (\mathfrak{P}_0, \mathfrak{L}_0)$ described above is uniquely reversible: given an affine plane $(\mathfrak{P}_0, \mathfrak{L}_0)$, one may uniquely reconstruct the missing line along with its points, thereby forming the **projective completion** $(\mathfrak{P}, \mathfrak{L})$ of the given affine plane. This process is described as follows:

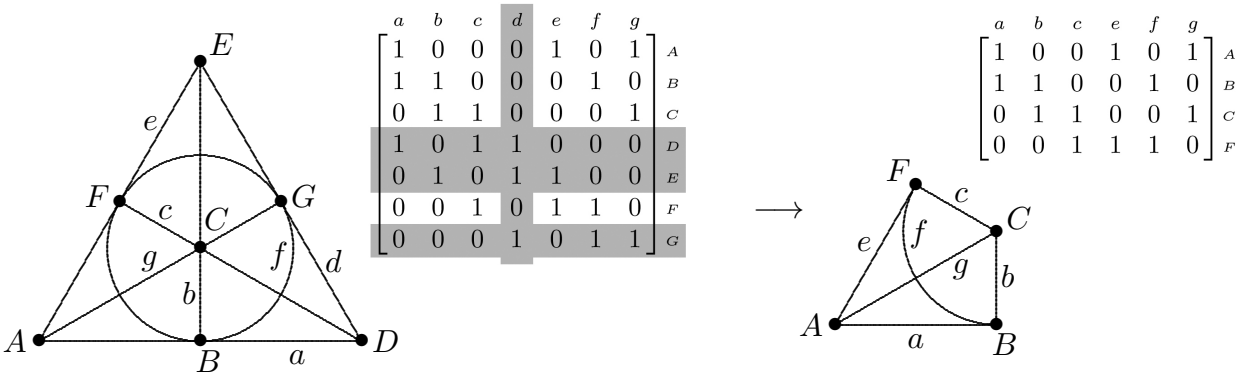
(1) For each parallel class of lines in $(\mathfrak{P}_0, \mathfrak{L}_0)$, add one new ideal point to all lines in the class.

(2) Add one new line whose points are the ideal points added in Step (1).

The new line added in Step (2) is sometimes called the *line at infinity* and may be denoted ℓ_∞. From the viewpoint of the original affine plane, this new line looks rather different from the other lines; but from the viewpoint of the projective plane $(\mathfrak{P}, \mathfrak{L})$, there is not typically any feature that distinguishes one line from another.

7.1 Example: The Projective Plane of Order 2. Let $(\mathfrak{P}, \mathfrak{L})$ be the projective plane of order 2, as shown on the left below. By deleting the line d and its three points

D, E, G we obtain the affine plane of order 2 with point set $\mathfrak{P}_0 = \{A, B, C, F\}$ and line set $\mathfrak{L}_0 = \{a, b, c, e, f, g\}$.



From the resulting affine plane we may uniquely reconstruct the original projective plane, as follows:

(1) Add the ideal point D to both lines of the parallel class $\{a, c\}$;
add the ideal point E to both lines of the parallel class $\{b, e\}$;
add the ideal point G to both lines of the parallel class $\{f, g\}$.

(2) Add the ideal line d whose points are just the ideal points D, E, G added in Step (1).

The choice of the line $d \in \mathfrak{P}$ was arbitrary; by deleting any of the seven lines, along with its points, we obtain an affine plane of order 2; and any of the resulting affine planes are in this case isomorphic.

7.2 Example: Classical Planes. Now consider a general field F. Let $(\mathfrak{P}, \mathfrak{L})$ be the classical plane $\mathbb{P}^2(F)$, and let ℓ_0 be a line. We may suppose, after a linear change of coordinates if necessary, that ℓ_0 is the line $\langle(0, 0, 1)^T\rangle$. Thus $\mathfrak{P}_0$ consists of all points $\langle(x, y, z)\rangle$ *not* lying on ℓ_0. All such points have $z \neq 0$, so we may multiply coordinates by z^{-1} to write every point of $\mathfrak{P}_0$ uniquely in the form $\langle(x, y, 1)\rangle$ where $x, y \in F$ are arbitrary. Every line $\ell \in \mathfrak{L}_0$ arises from a line $\langle(a, b, c)^T\rangle$ where $(a, b) \neq (0, 0)$ since $\ell \neq \ell_0$. If $b \neq 0$ then ℓ has the form $y = -(a/b)x - (c/b)$; otherwise $b = 0$ and $a \neq 0$ so ℓ has the form $x = -c/a$. Clearly $(\mathfrak{P}_0, \mathfrak{L}_0)$ is isomorphic to the classical affine plane $\mathbb{A}^2(F)$ under the map $\langle(x, y, 1)\rangle \mapsto (x, y)$.

The reverse process of reconstructing the projective completion of a given classical affine plane $\mathbb{A}^2(F)$ is also clear:

(1) For each $m \in F \cup \{\infty\}$, lines of slope m form a parallel class of lines in the affine plane $(\mathfrak{P}_0, \mathfrak{L}_0)$. Add a new ideal point to every line of such a parallel class. (Lines of slope ∞ are simply vertical lines of the form $x =$ constant.)

(2) Add an new line, whose points are just the ideal points named in Step (1).

It is worth emphasizing that when completing a given affine plane to a projective plane, each pair of parallel lines in the affine plane picks up just *one* ideal point of intersection in the projective completion. If one stares down a railroad track and imagines the two

parallel rails meeting 'at infinity', one might be tempted to associate two ideal points of intersection, one at each 'end', $+\infty$ and $-\infty$. This viewpoint misses the fact that the two endpoints of the projective line are effectively identified as one. In particular the real projective line is homeomorphic to a circle S^1. And for more general fields, which are not generally ordered, there is no distinction between positive and negative anyway; so there is no way to distinguish more than one sign of infinity.

We also emphasize that our process of completing the affine plane to a projective plane (whereby *many* new ideal points are added 'at infinity') is distinct from forming the 'one-point compactification' of a plane (in which just *one* new ideal point ∞ is introduced). Note that the one-point compactification of $\mathbb{R}^2$ is homeomorphic to the sphere S^2, often identified with the *Riemann sphere* $\mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$, e.g. via stereographic projection.

Consider the embedding of the classical affine plane $\mathbb{A}^2(F)$ in its projective completion $\mathbb{P}^2(F)$ as the complement of a line ℓ_0, which we may take to be the line $\langle (0,0,1)^T \rangle$; this embedding, as described above, maps points, lines, and more general algebraic curves, via

$$\begin{array}{rcl}
\mathbb{A}^2(F) & \hookrightarrow & \mathbb{P}^2(F) \smallsetminus \ell_0 \\
\text{point } (x,y) & \longrightarrow & \text{point } \langle (x,y,1) \rangle \\
\text{point } \left(\frac{X}{Z}, \frac{Y}{Z}\right) & \longleftarrow & \text{point } \langle (X,Y,Z) \rangle,\ Z \neq 0 \\
\left.\begin{array}{r}\text{curve } f(x,y)=0, \\ \deg f = k\end{array}\right\} & \xrightarrow{\text{homogenize}} & \text{curve } Z^k f\left(\frac{X}{Z}, \frac{Y}{Z}\right)=0 \\
\text{curve } g(x,y,1)=0 & \xleftarrow{\text{dehomogenize}} & \left\{\begin{array}{l}\text{curve } g(X,Y,Z)=0, \\ g(X,Y,Z) \text{ homogeneous}\end{array}\right.
\end{array}$$

For example consider the affine parabola $y = x^2$. Substituting $(x,y) = \left(\frac{X}{Z}, \frac{Y}{Z}\right)$, and then multiplying both sides by Z^2 to clear the denominators, the equation becomes $YZ = X^2$. The latter curve in $\mathbb{P}^2(F)$ has one more point than the original curve, namely the point $\langle (0,1,0) \rangle \in \ell_0$. Likewise the affine hyperbola $xy = 1$, when homogenized, becomes $XY = Z^2$; here the curve picks up two points at infinity, namely the points $\langle (1,0,0 \rangle, \langle (0,1,0) \rangle \in \ell_0$. Of course the two conics $YZ = X^2$ and $XY = Z^2$ are equivalent under a simple change of variable (interchanging X and Z). Thus the original two affine conics, one a parabola and the other a hyperbola, although not affinely equivalent, nevertheless are projectively equivalent regardless of the choice of field F.

Put another way, one may start with the conic $XY = Z^2$ in the projective plane, one may delete the line $\ell_0 : Z = 0$, a secant line meeting the conic in two points, $\langle (1,0,0) \rangle$ and $\langle (0,1,0) \rangle$, to obtain the affine hyperbola $xy = 1$ in the (x,y)-plane, where $(x,y) = \left(\frac{X}{Z}, \frac{Y}{Z}\right)$. Or one may delete the line $\ell_1 : X = 0$, a tangent line meeting the conic in just one point $\langle (0,1,0) \rangle$, to obtain the affine parabola $y = z^2$ in the (y,z)-plane, where $(y,z) = \left(\frac{Y}{X}, \frac{Z}{X}\right)$.

We return our consideration to the case of general (not necessarily classical) affine and projective planes. Every affine plane $(\mathfrak{P}_0, \mathfrak{L}_0)$ has a unique completion to a projective

plane $(\mathfrak{P}, \mathfrak{L})$ as described above. But given a projective plane $(\mathfrak{P}, \mathfrak{L})$, distinct choices of line $\ell_0, \ell_1 \in \mathfrak{L}$ yield affine parts (by the deletion of ℓ_0 or ℓ_1) which in general need not be isomorphic. If there exists $\sigma \in \mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$ such that $\ell_0^\sigma = \ell_1$, then σ induces an automorphism from the affine plane $(\mathfrak{P}\!\smallsetminus\![\ell_0], \mathfrak{L}\!\smallsetminus\!\{\ell_0\})$ to the affine plane $(\mathfrak{P}\!\smallsetminus\![\ell_1], \mathfrak{L}\!\smallsetminus\!\{\ell_1\})$. Because the projective plane $(\mathfrak{P}, \mathfrak{L})$ is uniquely reconstructible from either $(\mathfrak{P}\!\smallsetminus\![\ell_0], \mathfrak{L}\!\smallsetminus\!\{\ell_0\})$ or $(\mathfrak{P}\!\smallsetminus\![\ell_1], \mathfrak{L}\!\smallsetminus\!\{\ell_1\})$, the condition that these two affine planes are isomorphic forces ℓ_0 and ℓ_1 to be in the same orbit of $\mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$. Thus

7.3 Theorem. Let $\ell_0, \ell_1 \in \mathfrak{L}$ be lines of a projective plane $(\mathfrak{P}, \mathfrak{L})$. Then the affine planes $(\mathfrak{P}\!\smallsetminus\![\ell_0], \mathfrak{L}\!\smallsetminus\!\{\ell_0\})$ and $(\mathfrak{P}\!\smallsetminus\![\ell_1], \mathfrak{L}\!\smallsetminus\!\{\ell_1\})$ are isomorphic iff $\ell_0^\sigma = \ell_1$ for some $\sigma \in \mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$.

7.4 Example: Planes of Order 9. It is known [37] that there are exactly four isomorphism classes of projective planes of order 9: the classical plane $\mathbb{P}^2(\mathbb{F}_9)$, the (projective) Hall plane of order 9 (the projective completion of the affine translation plane described in Section 3.2), the dual of the Hall plane, and the Hughes plane of order 9. The number of line orbits in these planes is 1, 2, 2, 2 respectively. This gives a total of exactly $1 + 2 + 2 + 2 = 7$ isomorphism classes of affine planes of order 9.

Exercises 7.

1. The known projective planes of order 27 are listed as follows, along with the number of line orbits of the full automorphism group in each case:

No.	*Plane*	*Line Orbit Sizes*
(a)	$\mathbb{P}^2(\mathbb{F}_{27})$ (classical)	757
(b)	Generalized twisted field plane	1, 27, 729
(c)	Hering plane	1, 756
(c′)	dual of (c)	28,729
(d)	Flag-transitive affine plane (completed)	1, 756
(d′)	dual of (d)	28,729
(e)	Another flag-transitive affine plane (completed)	1, 756
(e′)	dual of (e)	28,729
(f)	Sherk plane	1, 27, 729
(f′)	dual of (f)	1, 27,729
(g)	André plane	1, 54, 702
(g′)	dual of (g)	2, 26,729
(h)	Figueroa plane	13, 312, 432

   Planes (a), (b), (h) are self-dual. How many isomorphism types of affine planes of order 27 are known?

2. Show that the cubic curves $y = x^3$ and $y^2 = x^3$ are projectively equivalent.

3. Show that the curves $y = x^2$ and $x^2 + y^2 = 1$ are projectively equivalent over $\mathbb{F}_3$ and over $\mathbb{F}_5$, but not over $\mathbb{F}_2$.

8. Advantages of the Projective Viewpoint

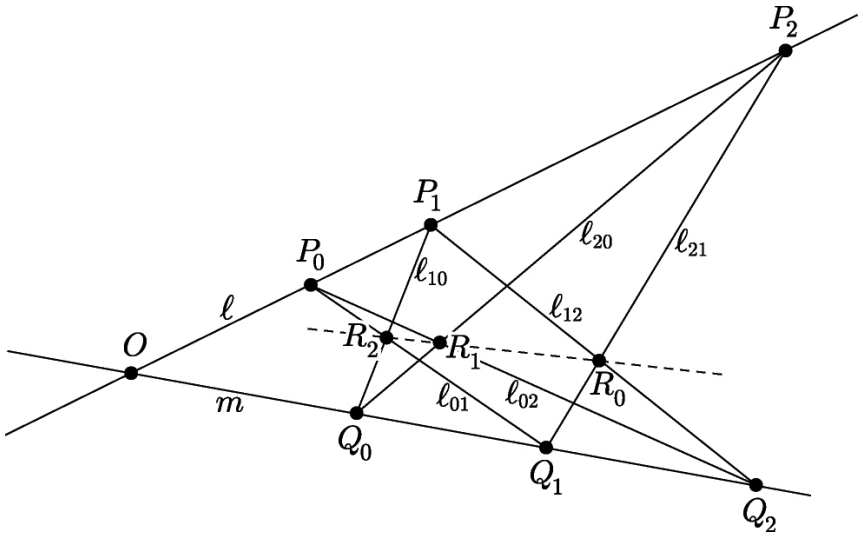
The axioms for projective planes are clearly simpler than those for affine planes. This suggests that for many purposes, the projective point of view is simpler than the affine point of view. We proceed to provide instances of this.

First of all, as we have seen, the theory of projective planes exhibits duality: the dual of every projective plane is also a projective plane. And because the dual of a classical plane is again classical, every theorem valid in the classical plane has a dual which is also valid in this plane. To illustrate the utility of this principle, consider the following result, known as the *Theorem of Pappus.* This basic result of classical plane geometry says that if

- ℓ and m are distinct lines in $\mathbb{P}^2(F)$, meeting in a point O;
- P_0, P_1, P_2 are distinct points of ℓ other than O;
- Q_0, Q_1, Q_2 are distinct points of m other than O;
- $\ell_{ij} = P_iQ_j$ for all $i \neq j$ in $\{0, 1, 2\}$;
- $R_i = \ell_{i+1,i+2} \cap \ell_{i+2,i+1}$ (with subscripts mod 3)

then the three points R_0, R_1, R_2 must be collinear.

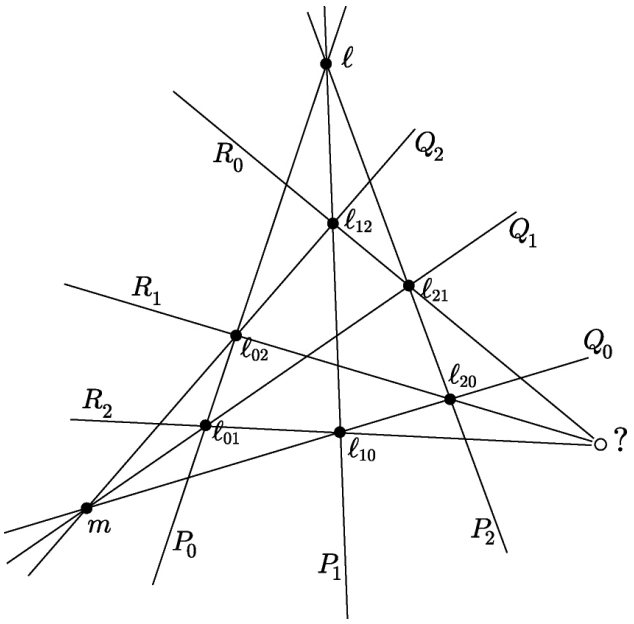
8.1 Figure.
Pappus' Configuration



This property does not hold in more general planes; a projective plane satisfies this condition iff it is isomorphic to $\mathbb{P}^2(F)$ for some field F. Later (Theorem 11.2) we will prove one direction (the 'only if' part) of this result. The dual of Pappus' Theorem, which also must hold in classical planes, states that if

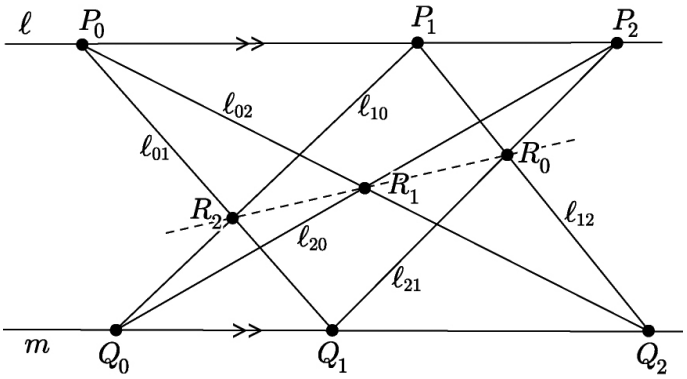
- ℓ and m are distinct points in $\mathbb{P}^2(F)$, joined by a line O;
- P_0, P_1, P_2 are distinct lines through ℓ other than O;
- Q_0, Q_1, Q_2 are distinct lines through m other than O;
- $\ell_{ij} = P_i \cap Q_j$ for all $i \neq j$ in $\{0, 1, 2\}$;
- $R_i = \ell_{i+1,i+2}\ell_{i+2,i+1}$ (with subscripts mod 3)

then the three lines R_0, R_1, R_2 must be concurrent.

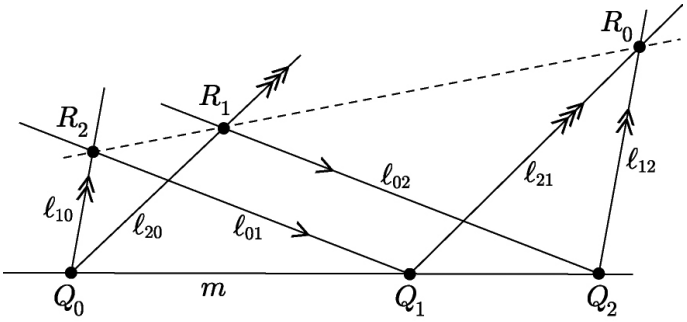


8.2 Figure.
Dual Pappus' Configuration

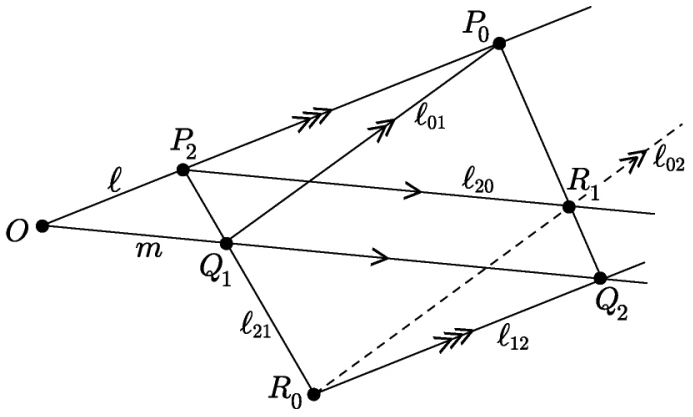
The Theorem of Pappus yields *many* theorem valid in the classical affine plane $\mathbb{A}^2(F)$, in particular the Euclidean plane $\mathbb{A}^2(\mathbb{R})$. One such theorem is illustrated by the figure



where we have chosen ℓ_∞ to be a line through O other than ℓ or m; the picture shows the affine plane formed by the complement of ℓ_∞. In the following figure we have designated pairs of parallel lines as usual with matching arrow markings. Here we choose ℓ itself to be the line ℓ_∞ at infinity:



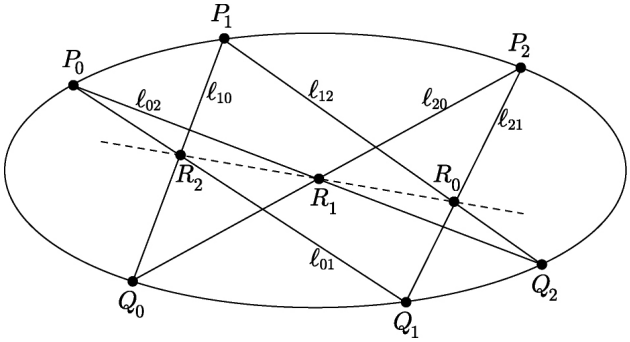
And in the following example, the line ℓ_{12} is taken as the line ℓ_∞ at infinity:



Of course the original version of Figure 8.1 (with no parallel lines) also holds in the classical affine plane. Further affine versions of this result may be listed; and then many affine versions of the dual of Pappus' Theorem may also be listed. *But all of these theorems valid in the classical affine plane are summarized a single result of classical projective geometry,* namely the original Theorem of Pappus.

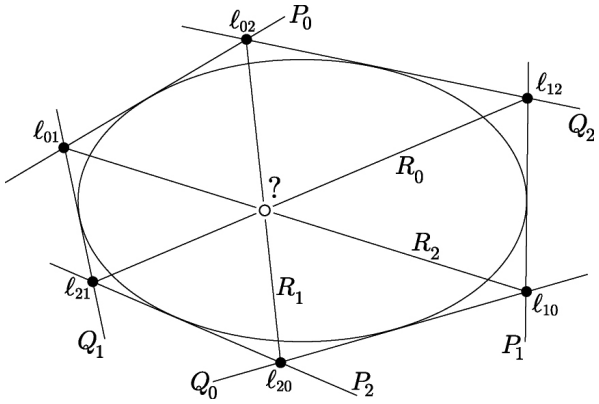
A generalization of Pappus' Theorem, known as **Pascal's Theorem**, holds in the classical projective plane $\mathbb{P}^2(F)$: here the pair of lines $\{\ell, m\}$ is replaced by a conic. The

8.3 Figure.
Pascal's Configuration



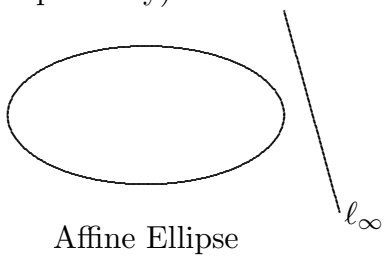
sense in which this generalizes Pappus' Theorem, is that a pair of lines may be seen as a degenerate conic. This theorem (whose proof we omit) has also a dual form, valid in the classical projective plane; to dualize, points of the conic are replaced by tangent lines, thus:

8.4 Figure.
Dual of Pascal's Configuration

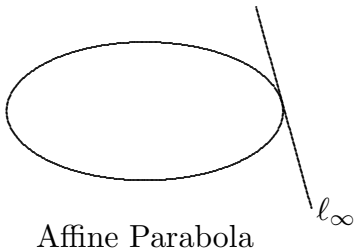


Both Figures 8.3 and 8.4 have multiple affine versions, as we saw with Pappus' Theorem; and all these versions are economically summarized in the single Theorem of Pascal. (In

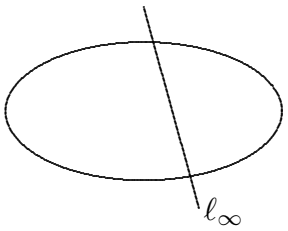
fact it is possible to give one proof that covers both Pascal's Theorem and Pappus' Theorem.) One of the reasons for the multiple affine versions of Pascal's Theorem is that in the affine plane there are three types of nondegenerate conics: ellipses (including circles), hyperbolas, and parabolas. If one were to prove Pascal's theorem in the affine setting, separate arguments would be required for all three cases. In the projective setting, all three cases are the same! This is because what appears to the affine observer as three separate types of conic (ellipses, parabolas, or hyperbolas) are simply the affine portions of a projective conic, formed by deleting some line ℓ_∞, which meets the conic in 0, 1 or 2 points respectively (i.e. ℓ_∞ is a passant line, a tangent line, or a secant line to the conic, respectively):



Affine Ellipse



Affine Parabola



Affine Hyperbola

In comparing the projective and affine viewpoints to plane geometry, we recognize several aspects of the projective viewpoint which recommend this as the easier, or more natural approach in many situations:

(1) The axioms for projective planes are much more concise than those for affine planes.

(2) The dual of a projective plane is simply a projective plane. The dual of an affine plane is not an affine plane; it is just ... well, it's a dual affine plane, OK?

(3) If one's goal is to classify planes, the list of projective planes of a given order is usually much simpler than the list of affine planes of that order. Consider for example the 7 known affine planes of order 9, which are equivalent to the list of 4 projective planes of order 9 up to isomorphism, or simply 3 projective planes up to isomorphism/duality (Example 7.4).

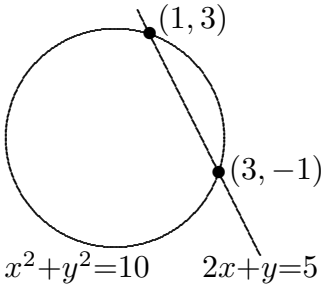
(4) It is possible for one object (such as a curve or point-line configuration) in the projective plane to appear as several distinct objects in its various affine sections. This means that one theorem in projective plane geometry can replace several theorems in the affine setting; or that one theorem can require a single proof in the projective case but many different proofs in the affine case.

(5) Some natural properties of the projective plane have no readily described affine counterpart.

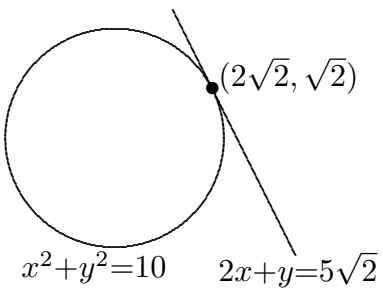
As an example of (5) we consider the problem of counting intersection points of plane curves. The curves we consider are algebraic curves, i.e. zero sets of polynomials. The degree of such a curve is the degree of its defining polynomial. As we have seen, we require homogeneous polynomials in order to obtain well-defined curves in the projective plane.

The following examples of curves are defined over the real field $\mathbb{R}$, although the same principles apply to arbitrary fields. We begin with the observation that a line (a curve of degree 1) generally meets a conic (a curve of degree 2) in two points; the case illustrated in Example 8.5 below is typical. In Example 8.6 the line is tangent to the circle and the point of intersection is a double point; so if points are counted with appropriate multiplicity, the number of intersection points is again 2. While it is possible to give a precise definition of intersection multiplicity, as is done in any course on algebraic curves, we will not do so here and now; we trust that a few examples will convey the general idea.

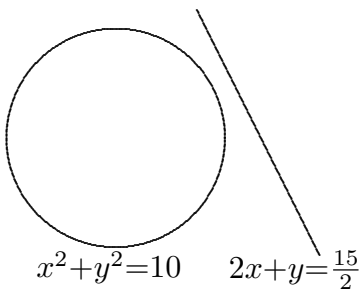
8.5 Example.



8.6 Example.



8.7 Example.



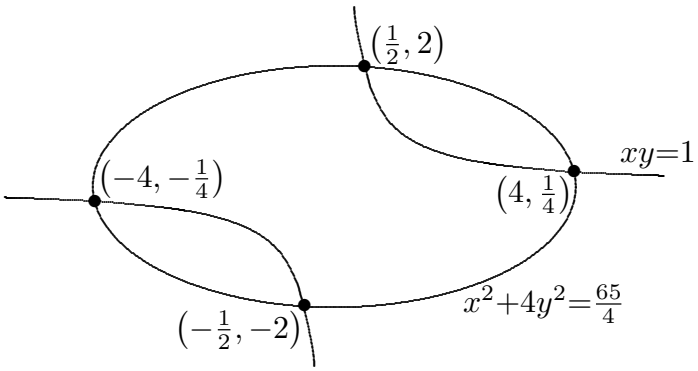
In Example 8.7 no intersection points of the indicated line and circle are visible; this is because the points of intersection (found algebraically) are $\left(3+\frac{1}{2}i, \frac{3}{2}-i\right)$ and $\left(3-\frac{1}{2}i, \frac{3}{2}+i\right)$. Although the curves themselves are *defined over* $\mathbb{R}$ (meaning that the defining polynomials have real coefficients) the actual *points* of the curve may have coordinates in any algebraic extension. In particular the set of all points of a curve is considered to be those points with coordinates in the algebraic closure of the field over which the curves are defined. All our examples will be defined over $\mathbb{R}$ and so their points will have complex coordinates.

Similarly when intersecting two conics we find typically 4 points of intersection, as observed below.

8.8 Example.

Four points of intersection.

Compare: $\deg(f)\deg(g) = 2{\cdot}2 = 4$

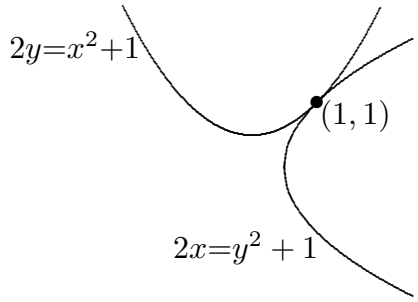


Why four points? Simply multiply the degrees of the two curves: $2 \times 2 = 4$. Again we must count points with appropriate multiplicity, and include all points with complex coordinates which simultaneously satisfy the two defining polynomial equations. Here for example we

intersect two parabolas and once again find 4 points of intersection:

8.11 Example.

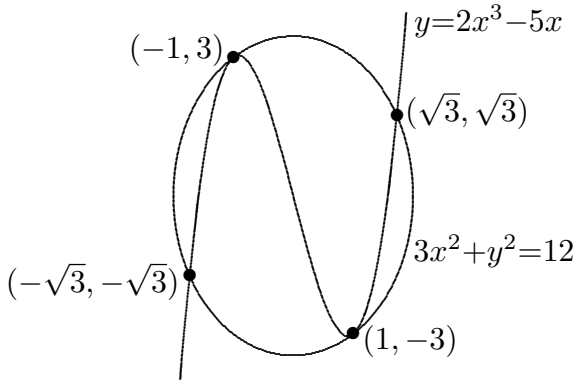
Four points of intersection:
$(1,1)$ (double point),
$(-1{+}2i, -1{-}2i)$, $(-1{-}2i, -1{+}2i)$.
Compare: $\deg(f)\deg(g) = 2{\cdot}2 = 4$



In the following example we intersect a conic with a cubic to obtain $2 \times 3 = 6$ points of intersection:

8.9 Example.

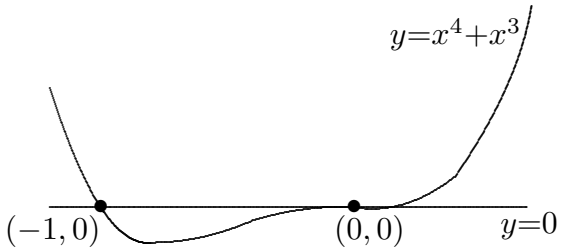
Six points of intersection:
$(-1,3)$, $(1,-3)$ (double points),
$(\sqrt{3},\sqrt{3})$, $(-\sqrt{3},-\sqrt{3})$.
Compare: $\deg(f)\deg(g) = 3{\cdot}2 = 6$



Similarly, intersecting a line with a quartic (a curve of degree 4) we expect $1 \times 4 = 4$ points of intersection. The following example realizes this total with one single point and one triple point:

8.10 Example.

Four points of intersection:
$(0,0)$ (triple point), $(-1,0)$.
Compare: $\deg(f)\deg(g) = 4{\cdot}1 = 4$



We have included this example so the student will see that what we mean (and have not defined) by intersection multiplicity, generalizes the usual notion of multiplicity of zeroes of a polynomial in one indeterminate: in this case $f(x) = x^4 + x^3 = x^3(x+1)$ has four zeroes counting multiplicity, with a triple zero at 0.

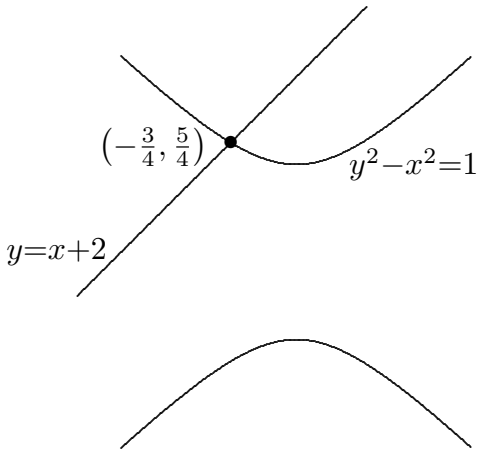
Now consider the following example, which at first seems to be a counterexample to the principle observed up to this point. We intersect a hyperbola with a line, and we expect to find $2 \times 1 = 2$ points of intersection. We see only one point of intersection (with

real or complex coordinates); where is the missing point? It is out there, at infinity!

8.12 Example.

Just one affine point of intersection: $\left(-\frac{3}{4}, \frac{5}{4}\right)$.

Compare: $\deg(f)\deg(g) = 2{\cdot}1 = 2$



The two curves have homogeneous equations $Y^2 - X^2 = Z^2$ and $Y = X + 2Z$. Now the two projective points of intersection are the points

$$\langle(-3, 5, 4)\rangle = \left\langle\left(-\tfrac{3}{4}, \tfrac{5}{4}, 1\right)\right\rangle \quad \text{and} \quad \langle(1, 1, 0)\rangle.$$

The second point lies on the line ℓ_∞, which is defined by $Z = 0$.

All these examples illustrate **Bezout's Theorem**, which says that given two plane curves of degree m and n respectively, assuming the two curves do not coincide or share any components, then the number of points of intersection equals mn. This elegant statement is valid if we abide by the following three principles:

- Points whose coordinates lie in the algebraic closure of the original field of definition, should be included.
- Points of intersection should be counted with appropriate multiplicity.
- The curves should be considered as curves in the *projective* plane, not just in the affine plane.

Exercises 8.

1. Consider the two Euclidean plane curves $y = x^2$ and $y = x^3$. Find all points of intersection in the complex projective plane. What are the corresponding intersection multiplicities?

9. Closed Configurations

We pause to consider a class of point-line incidence structures which slightly generalizes the notion of a projective plane, by allowing 'degenerate' cases which may fail the axiom (P3). Thus a **closed configuration** is an incidence system of points and lines such that

(P1) For any two distinct points, there is exactly one line through both.

(P2) Any two distinct lines meet in exactly one point.

A closed configuration satisfying also (P3) (i.e. containing at least four points of which no three are collinear) are of course projective planes; those which fail (P3) are said to be **degenerate**. Note that the point-line dual of a closed configuration, is again a closed configuration.

Closed configurations arise naturally as substructures of projective planes which are closed under the meet '$\wedge$' and join '$\vee$' operations, as any such substructure will inherit the properties (P1) and (P2) from the plane in which it is embedded. To appreciate the motivation for studying substructures which are closed under the operations of meet and join, the student is encouraged to recall that in the study of groups, one encounters very early the notion of a subgroup: a subset closed under the group operation, also satisfying the usual axioms of group theory. Same for subgraphs of graphs, subfields of fields, etc. In the case of projective planes, such closed substructures include not only subplanes (which are projective planes in their own right) but also degenerate closed configurations.

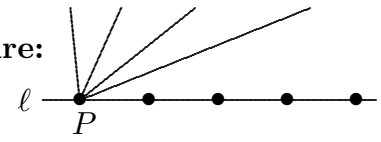
It is not hard to see that the following lists all possible closed configurations.

9.1 Proposition. A point-line incidence structure is a closed configuration iff it is one of the following:

(i) $(\emptyset, \emptyset)$;
(ii) a single point;
(iii) a single line;
(iv) a line ℓ and a point $P \in \ell$, together with $m \geq 0$ lines through P other than ℓ, and $n \geq 0$ points on ℓ other than P;
(v) a line ℓ and a point $P \notin \ell$, together with $m \geq 0$ lines through P and their m points of intersection with ℓ; or
(vi) a projective plane.

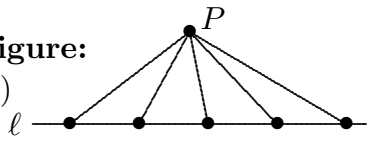
Examples typical of cases (iv) and (v) are shown:

9.1a Figure: Case (iv)



or

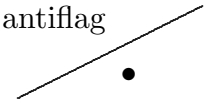
9.1b Figure: Case (v)



Two examples of closed configurations, which are the smallest cases of (iv) and (v) respectively, are the **flag** (an incident point-line pair) and **antiflag** (a non-incident point-line pair), as shown:

a flag

an antiflag



A typical projective plane contains many more degenerate closed configurations than subplanes. In finite classical planes $\mathbb{P}^2(F)$, every subfield $K \subseteq F$ gives rise to a subplane

isomorphic to $\mathbb{P}^2(K)$, given by those points and lines coordinatized by the subfield. In the classical case such subplanes, and their images under automorphisms of $\mathbb{P}^2(F)$, account for all subplanes. Thus for example in a finite classical projective plane $\mathbb{P}^2(\mathbb{F}_q)$, a subplane of order q_0 exists iff $q = q_0^r$ for some $r \geq 1$. For example we present an incidence matrix for the projective plane $\mathbb{P}^2(\mathbb{F}_4)$ of order 4, having 21 points and 21 lines. We have partitioned the incidence matrix so that three of its subplanes of order 2 are evident:

$$\begin{array}{ccc}
1\,1\,0\,1\,0\,0\,0 & 0\,1\,0\,0\,0\,0\,0 & 0\,0\,0\,1\,0\,0\,0 \\
0\,1\,1\,0\,1\,0\,0 & 0\,0\,1\,0\,0\,0\,0 & 0\,0\,0\,0\,1\,0\,0 \\
0\,0\,1\,1\,0\,1\,0 & 0\,0\,0\,1\,0\,0\,0 & 0\,0\,0\,0\,0\,1\,0 \\
0\,0\,0\,1\,1\,0\,1 & 0\,0\,0\,0\,1\,0\,0 & 0\,0\,0\,0\,0\,0\,1 \\
1\,0\,0\,0\,1\,1\,0 & 0\,0\,0\,0\,0\,1\,0 & 1\,0\,0\,0\,0\,0\,0 \\
0\,1\,0\,0\,0\,1\,1 & 0\,0\,0\,0\,0\,0\,1 & 0\,1\,0\,0\,0\,0\,0 \\
1\,0\,1\,0\,0\,0\,1 & 1\,0\,0\,0\,0\,0\,0 & 0\,0\,1\,0\,0\,0\,0 \\
\\
0\,0\,0\,0\,1\,0\,0 & 1\,1\,0\,1\,0\,0\,0 & 0\,1\,0\,0\,0\,0\,0 \\
0\,0\,0\,0\,0\,1\,0 & 0\,1\,1\,0\,1\,0\,0 & 0\,0\,1\,0\,0\,0\,0 \\
0\,0\,0\,0\,0\,0\,1 & 0\,0\,1\,1\,0\,1\,0 & 0\,0\,0\,1\,0\,0\,0 \\
1\,0\,0\,0\,0\,0\,0 & 0\,0\,0\,1\,1\,0\,1 & 0\,0\,0\,0\,1\,0\,0 \\
0\,1\,0\,0\,0\,0\,0 & 1\,0\,0\,0\,1\,1\,0 & 0\,0\,0\,0\,0\,1\,0 \\
0\,0\,1\,0\,0\,0\,0 & 0\,1\,0\,0\,0\,1\,1 & 0\,0\,0\,0\,0\,0\,1 \\
0\,0\,0\,1\,0\,0\,0 & 1\,0\,1\,0\,0\,0\,1 & 1\,0\,0\,0\,0\,0\,0 \\
\\
0\,0\,1\,0\,0\,0\,0 & 0\,0\,0\,0\,1\,0\,0 & 1\,1\,0\,1\,0\,0\,0 \\
0\,0\,0\,1\,0\,0\,0 & 0\,0\,0\,0\,0\,1\,0 & 0\,1\,1\,0\,1\,0\,0 \\
0\,0\,0\,0\,1\,0\,0 & 0\,0\,0\,0\,0\,0\,1 & 0\,0\,1\,1\,0\,1\,0 \\
0\,0\,0\,0\,0\,1\,0 & 1\,0\,0\,0\,0\,0\,0 & 0\,0\,0\,1\,1\,0\,1 \\
0\,0\,0\,0\,0\,0\,1 & 0\,1\,0\,0\,0\,0\,0 & 1\,0\,0\,0\,1\,1\,0 \\
1\,0\,0\,0\,0\,0\,0 & 0\,0\,1\,0\,0\,0\,0 & 0\,1\,0\,0\,0\,1\,1 \\
0\,1\,0\,0\,0\,0\,0 & 0\,0\,0\,1\,0\,0\,0 & 1\,0\,1\,0\,0\,0\,1
\end{array}$$

No such Lagrange-like condition holds for orders of subplanes in general; for example the Hughes plane of order 9 has a subplane of order 2. However the following bound holds.

9.2 Theorem. Let $(\mathfrak{P}, \mathfrak{L})$ be a plane of order n with a proper subplane $(\mathfrak{P}_0, \mathfrak{L}_0)$ of order n_0. Then $n_0^2 \leq n$, and equality holds iff every line $\ell \in \mathfrak{L}$ meets some point of $\mathfrak{P}_0$, iff every point $P \in \mathfrak{P}$ lies on some line of $\mathfrak{L}_0$.

Proof. Every $\ell \in \mathfrak{L}_0$ has $(n+1) - (n_0+1) = n - n_0$ points of $\mathfrak{P} \smallsetminus \mathfrak{P}_0$. Since no two lines of $\mathfrak{L}_0$ share any point of $\mathfrak{P} \smallsetminus \mathfrak{P}_0$, the total number of points covered by lines of $\mathfrak{L}_0$ is

$$\underbrace{(n_0^2 + n_0 + 1)}_{\text{points of } \mathfrak{P}_0} + \underbrace{(n_0^2 + n_0 + 1)(n - n_0)}_{\substack{\text{points of } \mathfrak{P} \smallsetminus \mathfrak{P}_0 \\ \text{covered by lines of } \mathfrak{L}_0}} \leq \underbrace{n^2 + n + 1}_{\text{points of } \mathfrak{P}},$$

which simplifies to

$$(n - n_0)(n - n_0^2) \geq 0.$$

Since $(\mathfrak{P}_0, \mathfrak{L}_0)$ is a proper subplane we have $n_0 < n$ and so $n_0^2 \leq n$. If equality holds here then equality must hold throughout, which means that every point of $\mathfrak{P}$ lies on some line of $\mathfrak{L}_0$. A dual argument (interchanging points and lines) shows that this condition is also equivalent to every line of $\mathfrak{L}$ meeting some point of $\mathfrak{P}_0$. □

A proper subplane $(\mathfrak{P}_0, \mathfrak{L}_0)$ of $(\mathfrak{P}, \mathfrak{L})$ is a **Baer subplane** if every line of the plane meets some point of the subplane $(\mathfrak{P}_0, \mathfrak{L}_0)$. The conclusion of Theorem 9.2 shows that in a plane of finite order n, a Baer subplane is simply a subplane of order $\sqrt{n}$; and this cannot exist unless n is a perfect square. By the remarks above, every finite classical projective plane of square order has Baer subplanes.

Exercises 9.

1. Let $(\mathfrak{P}, \mathfrak{L})$ be a projective plane of order n having a proper subplane $(\mathfrak{P}_0, \mathfrak{L}_0)$ of order n_0, so that $n_0^2 \leq n$ by Theorem 9.2. If $n_0^2 < n$, show that $n_0^2 + n_0 \leq n$. Say as much as you can about the case of equality: $n_0^2 + n_0 = n$ (giving examples if possible).

   Hint. Suppose $n_0^2 < n$, so there exists a line $\ell \in \mathfrak{L}$ not containing any point of $\mathfrak{P}_0$. Consider intersections of ℓ with lines of $\mathfrak{L}_0$.

2. We have shown that the points and lines of $\mathbb{P}^2(\mathbb{F}_4)$ can be partitioned into three subplanes of order 2. Can they also be partitioned into

   (a) 21 flags?

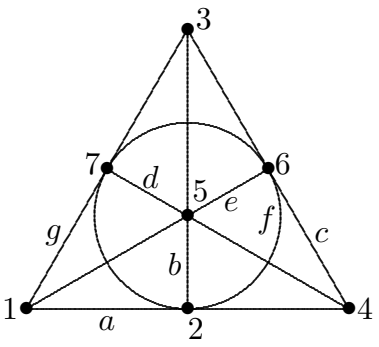
   (b) 21 antiflags?

   (c) 7 triangles ('subplanes of order 1')?

   Do similar partitions exist if $\mathbb{P}^2(\mathbb{F}_4)$ is replaced by a more general projective plane? Find such generalizations if possible.

10. Collineations and Correlations

Automorphisms of a projective plane $(\mathfrak{P}, \mathfrak{L})$ are often called **collineations**. (One reason for introducing the new term 'collineation' is to distinguish this type of automorphism from a 'correlation', which we define later in this Section.) Thus a collineation σ of $(\mathfrak{P}, \mathfrak{L})$ may be viewed as a permutation of $\mathfrak{P} \cup \mathfrak{L}$ mapping points to points, and lines to lines, such that $P^\sigma \in \ell^\sigma$ iff $P \in \ell$. The (full) automorphism group $\mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$ is the **(full) collineation group**, and subgroups of $\mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$ are **collineation groups**. For example consider the projective plane of order two, shown in Figure 10.1; here we have labelled the points $\mathfrak{P} = \{1, 2, \ldots, 7\}$ and the lines $\mathfrak{L} = \{a, b, \ldots, g\}$.



10.1 Figure
An Incidence Matrix A of the Plane $\mathbb{P}^2(\mathbb{F}_2)$

$$A = \begin{array}{c} \begin{matrix} a & b & c & d & e & f & g \end{matrix} \\ \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right] \end{array} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix}$$

The full collineation group of this plane has order 168, and is generated by the two collineations $(1234567)(abcdefg)$ and $(14)(67)(cg)(de)$.

Assume that the plane $(\mathfrak{P}, \mathfrak{L})$ has order n. Enumerate the points and lines as $\mathfrak{P} = \{P_1, P_2, \ldots, P_N\}$, $\mathfrak{L} = \{\ell_1, \ell_2, \ldots, \ell_N\}$ where $N = n^2 + n + 1$. Every collineation $\sigma \in \mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$ gives rise to a pair of permutation matrices $M(\sigma), M'(\sigma)$ representing the action of σ on points and lines, respectively; thus $M(\sigma) = \big(m_{ij}(\sigma)\big)_{1 \le i,j \le N}$ and $M'(\sigma) = \big(m'_{ij}(\sigma)\big)_{1 \le i,j \le N}$ where

$$m_{ij}(\sigma) = \begin{cases} 0, & \text{if } P_i^\sigma \neq P_j\,, \\ 1, & \text{if } P_i^\sigma = P_j\,; \end{cases} \qquad m'_{ij}(\sigma) = \begin{cases} 0, & \text{if } \ell_i^\sigma \neq \ell_j\,, \\ 1, & \text{if } \ell_i^\sigma = \ell_j\,. \end{cases}$$

Let $A = \big(a_{ij}\big)$ be the incidence matrix of A, so that

$$a_{ij} = \begin{cases} 1, & \text{if } P_i \in \ell_j; \\ 0, & \text{if } P_i \notin \ell_j. \end{cases}$$

Given two collineations $\sigma, \tau \in \mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$, the (i,j)-entry of $M(\sigma)M(\tau)$ is

$$\begin{aligned} \sum_k m_{ik}(\sigma) m_{kj}(\tau) &= \begin{cases} 1, & \text{if there exists } k \text{ such that } P_i \overset{\sigma}{\longmapsto} P_k \overset{\tau}{\longmapsto} P_j\,; \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} 1, & \text{if } P_i \overset{\sigma\tau}{\longmapsto} P_j\,; \\ 0, & \text{otherwise} \end{cases} \\ &= m_{ij}(\sigma\tau) \end{aligned}$$

which shows that $M(\sigma)M(\tau) = M(\sigma\tau)$, i.e. $M : \mathrm{Aut}(\mathfrak{P}, \mathfrak{L}) \to GL_N(\mathbb{Q})$ is a homomorphism. Note also that

$$M(\sigma)^T = M(\sigma)^{-1} = M(\sigma^{-1})$$

since $M(\sigma)$ is a permutation matrix, hence orthogonal. Similar properties hold for $M' : \mathrm{Aut}(\mathfrak{P}, \mathfrak{L}) \to GL_N(\mathbb{Q})$ in place of M. Now if $\sigma \in \mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$ then the (i,j)-entry of $M(\sigma)A$ is

$$\begin{aligned} \sum_k m_{ik}(\sigma) a_{kj} &= \begin{cases} 1, & \text{if there exists } k \text{ such that } P_i \overset{\sigma}{\longmapsto} P_k \in \ell_j\,; \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} 1, & \text{if } P_i^\sigma \in \ell_j\,; \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

whereas the (i,j)-entry of $AM'(\sigma)$ is

$$\begin{aligned} \sum_k a_{ik} m_{kj}(\sigma) &= \begin{cases} 1, & \text{if there exists } k \text{ such that } P_i \in \ell_k \overset{\sigma}{\longmapsto} \ell_j\,; \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} 1, & \text{if } P_i^\sigma \in \ell_j\,; \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

and so $M(\sigma)A = AM'(\sigma)$. Conversely if M_1, M_2 are $N \times N$ permutation matrices satisfying $M_1 A M_2^{-1} = A$, then the corresponding pair of permutations of $\mathfrak{P}$ and $\mathfrak{L}$ is a collineation of the plane $(\mathfrak{P}, \mathfrak{L})$. This proves

10.2 Theorem. $\mathrm{Aut}(\mathfrak{P}, \mathfrak{L}) \cong \{\text{pairs } (M_1, M_2) \text{ of } N \times N \text{ permutation matrices } : M_1 A = A M_2\}$.

For example in the case of the projective plane of order two listed above, the generators for the collineation group listed above are represented in matrix form as

$$\underbrace{\begin{bmatrix} 0&1&0&0&0&0&0\\ 0&0&1&0&0&0&0\\ 0&0&0&1&0&0&0\\ 0&0&0&0&1&0&0\\ 0&0&0&0&0&1&0\\ 0&0&0&0&0&0&1\\ 1&0&0&0&0&0&0 \end{bmatrix}}_{(1234567)} \underbrace{\begin{bmatrix} 1&0&0&0&1&0&1\\ 1&1&0&0&0&1&0\\ 0&1&1&0&0&0&1\\ 1&0&1&1&0&0&0\\ 0&1&0&1&1&0&0\\ 0&0&1&0&1&1&0\\ 0&0&0&1&0&1&1 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1&0&0&0&1&0&1\\ 1&1&0&0&0&1&0\\ 0&1&1&0&0&0&1\\ 1&0&1&1&0&0&0\\ 0&1&0&1&1&0&0\\ 0&0&1&0&1&1&0\\ 0&0&0&1&0&1&1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 0&1&0&0&0&0&0\\ 0&0&1&0&0&0&0\\ 0&0&0&1&0&0&0\\ 0&0&0&0&1&0&0\\ 0&0&0&0&0&1&0\\ 0&0&0&0&0&0&1\\ 1&0&0&0&0&0&0 \end{bmatrix}}_{(abcdefg)};$$

$$\underbrace{\begin{bmatrix} 0&0&0&1&0&0&0\\ 0&1&0&0&0&0&0\\ 0&0&1&0&0&0&0\\ 1&0&0&0&0&0&0\\ 0&0&0&0&1&0&0\\ 0&0&0&0&0&0&1\\ 0&0&0&0&0&1&0 \end{bmatrix}}_{(14)(67)} \underbrace{\begin{bmatrix} 1&0&0&0&1&0&1\\ 1&1&0&0&0&1&0\\ 0&1&1&0&0&0&1\\ 1&0&1&1&0&0&0\\ 0&1&0&1&1&0&0\\ 0&0&1&0&1&1&0\\ 0&0&0&1&0&1&1 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1&0&0&0&1&0&1\\ 1&1&0&0&0&1&0\\ 0&1&1&0&0&0&1\\ 1&0&1&1&0&0&0\\ 0&1&0&1&1&0&0\\ 0&0&1&0&1&1&0\\ 0&0&0&1&0&1&1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 1&0&0&0&0&0&0\\ 0&1&0&0&0&0&0\\ 0&0&0&0&0&0&1\\ 0&0&0&0&1&0&0\\ 0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0\\ 0&0&1&0&0&0&0 \end{bmatrix}}_{(cg)(de)}.$$

A consequence of Theorem 10.2 is

10.3 Corollary. Every collineation of a plane $(\mathfrak{P}, \mathfrak{L})$ has equally many fixed points and fixed lines.

Proof. Let $\sigma \in \mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$. Note that the trace of $M_1(\sigma)$ is

$$\sum_i m_{ii}(\sigma) = \big|\{i \in \{1, 2, \ldots, N\} \,:\, P_i^\sigma = P_i\}\big|$$

is simply the number of points $P \in \mathfrak{P}$ fixed by σ. Similarly the trace of $M_2(\sigma)$ is the number of fixed lines. However the incidence matrix A is invertible:

$$A\big(\tfrac{1}{n}A^T - \tfrac{1}{n(n+1)}J\big) = \tfrac{1}{n}(nI + J) - \tfrac{1}{n(n+1)}(n+1)J = I.$$

Thus the matrices $M_1(\sigma)$ and $M_2(\sigma) = A^{-1}M_1(\sigma)A$ are similar. Since similar matrices have the same trace, the result follows. □

The argument used in our proof of Corollary 10.3 is based on concepts from representation theory; see Appendix A3. Here we have a pair of homomorphisms $\pi_i : \mathrm{Aut}(\mathfrak{P}, \mathfrak{L}) \to GL_N(\mathbb{Q})$, $\sigma \mapsto M_i(\sigma)$ for $i = 1, 2$. Since the matrix A is invertible, the representations π_1 and π_2 are *equivalent*. The same technique applies to any invertible matrix A, where we consider an automorphism of A as a pair of permutations of the rows and the columns; the argument shows that any such automorphism fixes equally many rows and columns. A more general formulation of this fact is given by

10.4 Theorem. Every group G of automorphisms of a finite projective plane $(\mathfrak{P}, \mathfrak{L})$ has equally many orbits on points and on lines.

Before proving Theorem 10.4, consider the following example: Let $(\mathfrak{P}, \mathfrak{L})$ be the projective plane of order two shown in Figure 10.1, and let $G = \langle g, h \rangle$ be the group of order four generated by $g = (12)(57)(bg)(ef)$ and $h = (15)(27)(ad)(bg)$. Then G has four point orbits $\{1, 2, 5, 7\}, \{3\}, \{4\}, \{6\}$ and four line orbits $\{a, d\}, \{b, g\}, \{c\}, \{e, f\}$ in accordance with Theorem 10.4. Note that the point orbits and line orbits do not have the same size! Moreover there is no canonical bijection between point and line orbits. Nevertheless every nonidentity element of G has the same number of fixed points and fixed lines, in this case three.

Proof of Theorem 10.4. For every point $P \in \mathfrak{P}$, denote the corresponding stabilizer $G_P = \{g \in G : P^g = P\}$ and orbit $P^G = \{P^g : g \in G\}$. Recall that $|P^G| = [G : G_P]$; see Theorem A2.2. Let $P_1, \ldots, P_r \in \mathfrak{P}$ be representatives of the distinct point orbits, and $\ell_1, \ldots, \ell_s \in \mathfrak{L}$ representatives of the distinct line orbits. We count

$$\begin{aligned}
\sum_{g \in G} trM_1(g) &= \sum_g \big|\{P : P^g = P\}\big| \\
&= \big|\{(P, g) \in \mathfrak{P} \times G : P^g = P\}\big| \\
&= \sum_{1 \le i \le r} |P_i^G||G_{P_i}| \\
&= \sum_{1 \le i \le r} |G| \\
&= r|G|.
\end{aligned}$$

Since $trM_1(g) = trM_2(g)$ we obtain

$$r|G| = \sum_{g \in G} trM_1(g) = \sum_{g \in G} trM_2(g) = s|G|. \qquad \square$$

Let σ be a collineation of a finite projective plane $(\mathfrak{P}, \mathfrak{L})$, and let $(\mathfrak{P}_\sigma, \mathfrak{L}_\sigma)$ be the substructure formed by the fixed points and lines; thus

$$\mathfrak{P}_\sigma = \{P \in \mathfrak{P} \,:\, P^\sigma = P\} \quad \text{and} \quad \mathfrak{L}_\sigma = \{\ell \in \mathfrak{L} \,:\, \ell^\sigma = \ell\}.$$

It is easy to see that this is a closed substructure of $(\mathfrak{P}, \mathfrak{L})$: if $P, Q \in \mathfrak{P}_\sigma$ then

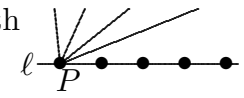
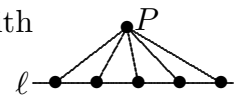
$$(PQ)^\sigma = P^\sigma Q^\sigma = PQ$$

and if $\ell, m \in \mathfrak{L}_\sigma$ then

$$(\ell \cap m)^\sigma = \ell^\sigma \cap m^\sigma = \ell \cap m.$$

The possibilities for a closed configuration were listed in Proposition 9.1; of these, Corollary 10.3 requires that we select only those with equally many points and lines as possible fixed substructures of collineations. This yields

10.5 Theorem. Let σ be a collineation of a finite projective plane $(\mathfrak{P}, \mathfrak{L})$. Then the fixed substructure of σ is a closed configuration of one of the types listed in the following table:

case	description of σ	$(\mathfrak{P}_\sigma, \mathfrak{L}_\sigma)$
(i)	**fixed-point free**	$(\emptyset, \emptyset)$
(ii)	**generalized (P,ℓ)-elation**	a line ℓ and a point $P \in \ell$, together with $m \geq 0$ lines through P other than ℓ, and $m \geq 0$ points on ℓ other than P 
(iii)	**generalized (P,ℓ)-homology**	a line ℓ and a point $P \notin \ell$, together with $m \geq 0$ lines through P and their m points of intersection with ℓ 
(iv)	**planar**	subplane

Technically, the identity is planar, since its fixed substructure is the entire plane. In cases (ii) and (iii), σ is called a **generalized (P,ℓ)-perspectivity**. As special cases, we call σ a **(P,ℓ)-perspectivity** if $m = n$, the order of the plane. This means that σ fixes every point of ℓ and every line through P; then we call σ a **(P,ℓ)-elation** if $P \in \ell$, or a **(P,ℓ)-homology** if $P \notin \ell$. Also in such cases we call P the **centre** and ℓ the **axis** of the perspectivity σ. For a fixed group of automorphisms $G \leq \mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$ and point-line pair (P, ℓ), the set of all $g \in G$ fixing every point of ℓ and every line through P forms a subgroup denoted $G(P, \ell) \leq G$; this group consists of the identity and all (P,ℓ)-perspectivities in G (if any).

Another special case of (ii) and (iii) is where $m = 0$: in this case $(P_\sigma, \mathfrak{L}_\sigma) = (\{P\}, \{\ell\})$. Here we say σ is a **flag collineation** or an **antiflag collineation**, according as $P \in \ell$ or $P \notin \ell$. Also case (iii) includes the case $m{=}2$ where σ is **triangular**: its fixed substructure is a triangle. In this case any of the three vertices of the triangle can be considered the centre of σ. Perhaps this should be considered a special case of (iv) really (a 'subplane of order 1', i.e. a 'thin' subplane; see Exercise #6.3).

More generally, every collineation group G has a fixed substructure $(\mathfrak{P}_G, \mathfrak{L}_G)$ which is a closed configuration in the plane $(\mathfrak{P}, \mathfrak{L})$. However, such closed configurations need not have

equally many points and lines (see the example given after the statement of Theorem 10.4 above, with three fixed points and only one fixed line). More general collineation groups can have fixed substructures of any of the types listed in Proposition 9.1.

As first examples illustrating Theorem 10.5, we classify some collineations of the projective plane of order 2 illustrated in Figure 10.1:

$(37)(56)(bf)(cd)$ is an $(1, a)$-elation;
$(24)(3576)(bdfc)(eg)$ is a flag collineation with centre 1 and axis a;
$(235)(476)(age)(cdf)$ is an antiflag collineation with centre 1 and axis b;
$(1234567)(abcdefg)$ is a fixed-point-free collineation.

This plane has no nontrivial homologies or planar or triangular collineations.

Consider now the classical plane $\mathbb{P}^2(F)$ over a field F. Here the full collineation group is $G = P\Gamma L_3(F)$, the group generated by all semilinear transformations $F^3 \to F^3$, i.e. transformations of the form $v \mapsto v^\alpha A$ where $A \in GL_3(F)$ and $\alpha \in \operatorname{Aut} F$; here we view F^3 as consisting of row vectors. (See Appendix A2.) In the group $P\Gamma L_3(F)$, all scalar transformations $v \mapsto \lambda v$ where $0 \neq \lambda \in F$ fix every point and every line of $\mathbb{P}^2(F)$ and so are the identity collineation. The elation group $G(\langle(1,0,0)\rangle, \langle(0,1,0)^T\rangle)$ consists of the transformations

$$\begin{bmatrix} 1 & c & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{i.e. } (x,\, y,\, z) \mapsto (x,\, y{+}cx,\, z).$$

The homology group $G(\langle(1,0,0)\rangle, \langle(0,1,0)^T\rangle)$ consists of the transformations

$$\begin{bmatrix} c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{i.e. } (x,\, y,\, z) \mapsto (cx,\, y,\, z)$$

where $c \neq 0$. Note that the elation and homology groups with fixed axis and centre are abelian, in fact isomorphic to the additive group of F and the multiplicative group $F^\times$ respectively. Nonclassical planes can however have nonabelian groups of perspectivities.

Generalizing the previous two examples, it is the case that invertible matrices of the form $I + B$ where B has rank one, represent perspectivities. The centre and axis are the row and column spaces of B, respectively. (Reverse these, if the underlying vector space F^3 is considered as column vectors.) Every nontrivial perspectivity of $\mathbb{P}^2(F)$ has this form.

Diagonalizable matrices with three distinct eigenvalues, represent triangular collineations (and conversely). For example the diagonal matrix

$$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

has fixed substructure consisting of the triangle with vertices

$$\langle(1,0,0)\rangle, \quad \langle(0,1,0)\rangle, \quad \langle(0,0,1)\rangle.$$

More generally a diagonalizable matrix with three distinct eigenvalues, has three 1-dimensional eigenspaces, and these are the vertices of its fixed triangle.

A matrix of the form

$$A = \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$$

with $ab \neq 0$ has a unique eigenspace $\langle (0,0,1) \rangle$ which is the centre of the corresponding collineation; its axis is the line $x = 0$, i.e. $\langle (1,0,0)^T \rangle$ so A represents a flag collineation.

A matrix of the form

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{bmatrix},$$

assuming the 2×2 matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ has an irreducible characteristic polynomial, has $\langle (1,0,0) \rangle$ as its unique eigenspace. In this case A represents an antiflag collineation with centre $\langle (1,0,0) \rangle$ and axis $\langle (1,0,0)^T \rangle$.

A field automorphism $\alpha \in \operatorname{Aut} F$ acts on $\mathbb{P}^2(F)$ as $(x,y,z) \mapsto (x^\alpha, y^\alpha, z^\alpha)$. The corresponding collineation is planar with fixed substructure $\mathbb{P}^2(K)$ where $K = \{x \in F : x^\alpha = x\}$ is the fixed subfield of α.

10.6 Proposition. Let $(\mathfrak{P}, \mathfrak{L})$ be a projective plane of finite order n, and consider any group of automorphisms $G \leq \operatorname{Aut}(\mathfrak{P}, \mathfrak{L})$. Then $|G(P,\ell)|$ divides n or $n-1$ according as $P \in \ell$ or $P \notin \ell$.

Proof. Let m be a line through P other than ℓ. Let S be the set of points of m other than P, and not on ℓ. Then $G(P,\ell)$ permutes S. If $\sigma \in G(P,\ell)$ fixes a point of S, as well as fixing every point of m and every line through P, then by Theorem 10.5 we must have $\sigma = 1$. Thus every orbit of $G(P,\ell)$ on S has size $|G(P,\ell)|$ by Theorem A2.2. Since $|S| = n$ or $n-1$ according as $P \in \ell$ or $P \notin \ell$, the result follows. □

From the explicit description of the groups $G(P,\ell)$ above in the classical case, we see that the upper bound of Proposition 10.6 is attained in this case. Thus in the classical plane $\mathbb{P}^2(F)$ with full collineation group $G = P\Gamma L_3(F)$:

- If $P \in \ell$ and m is any line through P other than ℓ, then the group $G(P,\ell)$ of (P,ℓ)-elations permutes the points of $m \setminus \{P\}$ regularly. The group $G(P,\ell)$ is isomorphic to the additive group of F. In particular if $F = \mathbb{F}_q$ then $G(P,\ell)$ is elementary abelian of order q.
- If $P \notin \ell$ and m is any line through P other than ℓ, then the group $G(P,\ell)$ of (P,ℓ)-homologies permutes the points of $m \setminus \{P, \ell \cap m\}$ regularly. The group

$G(P,\ell)$ is isomorphic to the multiplicative group $F^\times$. In particular if $F = \mathbb{F}_q$ then $G(P,\ell)$ is cyclic of order $q-1$.

An element $\tau \in \mathrm{Aut}(\mathfrak{P},\mathfrak{L})$ of order two is called an **involutory collineation**, or simply an **involution**.

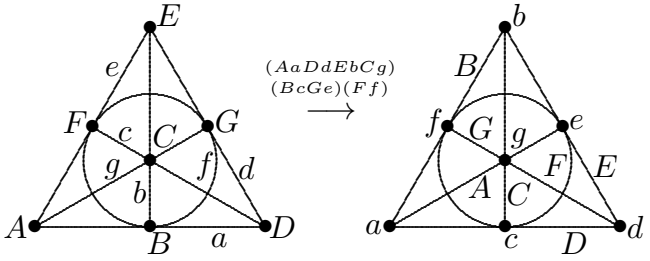
10.7 Theorem. Let τ be an involutory collineation of a projective plane $(\mathfrak{P},\mathfrak{L})$. Then τ is one of the following:

(a) an elation;
(b) a homology; or
(c) Baer planar.

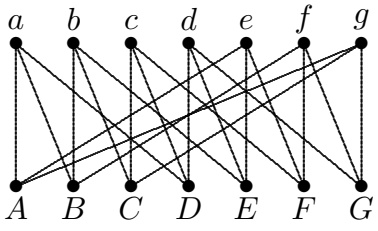
If the plane $(\mathfrak{P},\mathfrak{L})$ is of finite order n then cases (a), (b), (c) require that n is even, odd, or a perfect square, respectively.

Proof. It is not hard to see that every point P lies on some fixed line. If $\ell \in \mathfrak{L}$ is not fixed by τ, then $(\ell \cap \ell^\tau)^\tau = \ell^\tau \cap \ell = \ell \cap \ell^\tau$. In particular every non-fixed line passes through a fixed point. From the list of possible fixed substructures (i)–(v) above, we see that the fixed points consist either of all points of some line, or a subplane (which by Theorem 9.2 is necessarily Baer). Thus every involution is either an elation, a homology or Baer planar. If τ is a perspectivity then $G(P,\ell)$ contains an element of order 2 and so the parity of n follows from Proposition 10.6. □

A **correlation** of a projective plane $(\mathfrak{P},\mathfrak{L})$ is a permutation of $\mathfrak{P}\cup\mathfrak{L}$ which maps points to lines, and lines to points (thus interchanging the sets $\mathfrak{P}$ and $\mathfrak{L}$) while preserving (or, more correctly, reversing) incidence. So a correlation σ has the property that $\ell^\sigma \in P^\sigma$ iff $P \in \ell$. Such a correlation σ can be seen as an isomorphism between the original plane $(\mathfrak{P},\mathfrak{L})$ and the dual plane $(\mathfrak{L},\mathfrak{P})$, so only a self-dual plane can have any correlations at all. For example we illustrate here a correlation $(AaDdEbCg)(BcGe)(Ff)$ of the projective plane of order two:



10.8 Figure Incidence Graph of $\mathbb{P}^2(\mathbb{F}_2)$



We may view both collineations and correlations as automorphisms of the **incidence graph** of the projective plane $(\mathfrak{P},\mathfrak{L})$; this is the bipartite graph $\Gamma = \Gamma(\mathfrak{P},\mathfrak{L})$ having vertex set $\mathfrak{P}\cup\mathfrak{L}$, and with an edge from vertex $P \in \mathfrak{P}$ to vertex $\ell \in \mathfrak{L}$ iff the point P lies on the line ℓ. In fact the automorphisms of the graph Γ are of two types:

(i) Those $\sigma \in \operatorname{Aut}\Gamma$ preserving both $\mathfrak{P}$ and $\mathfrak{L}$. These correspond to collineations of the plane $(\mathfrak{P}, \mathfrak{L})$; and

(ii) Those $\sigma \in \operatorname{Aut}\Gamma$ interchanging $\mathfrak{P}$ and $\mathfrak{L}$, if they exist. These correspond to correlations of the plane $(\mathfrak{P}, \mathfrak{L})$.

So either the plane $(\mathfrak{P}, \mathfrak{L})$ is self-dual, in which case $\operatorname{Aut}(\mathfrak{P}, \mathfrak{L})$ (the collineation group) is a subgroup of index 2 in $\operatorname{Aut}\Gamma$ (identified with the group of all collineations and correlations of $(\mathfrak{P}, \mathfrak{L})$), or the plane $(\mathfrak{P}, \mathfrak{L})$ is not self-dual, in which case $\operatorname{Aut}\Gamma$ may be identified with the collineation group $\operatorname{Aut}(\mathfrak{P}, \mathfrak{L})$.

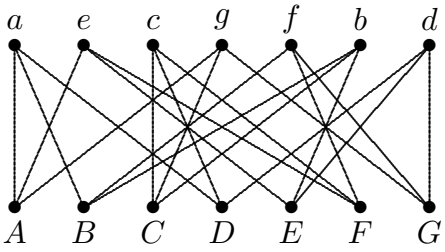
Figure 10.8 shows the projective plane of order two, along with its incidence graph, which evidently has automorphism group of order $2 \times 168 = 336$.

Note that every correlation of $(\mathfrak{P}, \mathfrak{L})$ must have even order. A **polarity** of $(\mathfrak{P}, \mathfrak{L})$ is a correlation of order 2. Given a polarity σ and an enumeration of points as $\mathfrak{P} = \{P_1, P_2, \ldots, P_N\}$, it is natural to list lines in the corresponding order $\mathfrak{L} = \{P_1^\sigma, P_2^\sigma, \ldots, P_N^\sigma\}$; in this case the incidence matrix is symmetric since $P_i \in P_j^\sigma$ iff $P_j \in P_i^\sigma$. (Here we have used the fact that $\sigma^2 = 1$ in order to simplify $(P_j^\sigma)^\sigma = P_j$.) Conversely, if a projective plane has a symmetric incidence matrix for some enumeration of its points P_i and lines ℓ_j, then the projective plane has a polarity $P_i \leftrightarrow \ell_i$.

For example the projective plane of order two illustrated in Figure 10.8 has a polarity $(Aa)(Be)(Cc)(Dg)(Ef)(Fb)(Gd)$, giving rise to the following symmetric incidence matrix.

10.9 Figure
The Polarity
$(Aa)(Be)(Cc)(Dg)(Ef)(Fb)(Gd)$

	a	e	c	g	f	b	d
A	1	1	0	1	0	0	0
B	1	0	0	0	1	1	0
C	0	0	1	1	0	1	0
D	1	0	1	0	0	0	1
E	0	1	0	0	0	1	1
F	0	1	1	0	1	0	0
G	0	0	0	1	1	0	1



Let σ be a polarity of a projective plane $(\mathfrak{P}, \mathfrak{L})$. A point P is **absolute** if $P^\sigma \in P$; otherwise P is **nonabsolute**. Similarly a line ℓ is absolute if $\ell^\sigma \in \ell$; otherwise ℓ is nonabsolute. The absolute points and lines are identified by 1's on the diagonal of the symmetric incidence matrix corresponding to σ. For the example illustrated in Figure 10.9, the absolute points are A, C, G and the absolute lines are a, c, d. Since the accompanying incidence graph lists points and lines in consistent order (corresponding via σ), the graph is symmetric about a horizontal axis of symmetry and the vertical edges identify the absolute points and lines. Note that the absolute points are collinear, and the absolute lines are concurrent. In general we have

10.10 Theorem. Suppose σ is a polarity of a projective plane $(\mathfrak{P}, \mathfrak{L})$ of order n, where n is not a square. Then σ has $n+1$ absolute points. These points form a line if n is even, or an oval if n is odd.

For a proof, see [31, Thm 12.7]. For a finite classical plane $\mathbb{P}^2(\mathbb{F}_q)$, one possible polarity is given by the transpose map which interchanges row and column vectors. Clearly, a point $\langle(x,y,z)\rangle$ is absolute iff

$$0 = (x,y,z)(x,y,z)^T = x^2 + y^2 + z^2.$$

Such points form a nondegenerate conic if q is odd, or a line $x+y+z=0$ if q is even. For classical planes of non-square order, every polarity has this form (called an **orthogonal polarity**) for some choice of coordinates. But suppose $q = q_0^2$, so that the quadratic extension $\mathbb{F}_q \supset \mathbb{F}_{q_0}$ has an automorphism $x \mapsto \overline{x} = x^{q_0}$ of order two. In this case the conjugate-transpose map $v \mapsto \overline{v}^T$ interchanges row and column vectors of length 3 and gives a polarity of $\mathbb{P}^2(\mathbb{F}_q)$, called a **unitary polarity**. The absolute points of this polarity are those satisfying the condition

$$0 = (x,y,z)(\overline{x},\overline{y},\overline{z})^T = x^{q_0+1} + y^{q_0+1} + z^{q_0+1}.$$

One checks in this case that there are q_0^3+1 absolute points (and lines), and $q_0^2(q_0^2-q_0+1)$ nonabsolute points (and lines). Moreover, the absolute points and nonabsolute lines form a 2-$(q_0^3+1,\, q_0+1,\, 1)$ design embedded in $(\mathfrak{P},\mathfrak{L})$, called a **Hermitian unital of order q_0.** We will encounter this example again in Example 18.7. Every polarity of $\mathbb{P}^2(\mathbb{F}_q)$ is either an orthogonal polarity (for arbitrary q) or a unitary polarity (for q a square only). Such polarities are as described above, after a change of coordinates of necessary.

Exercises 10.

1. For each of the planes indicated, find the number of collineations of each of the types listed in Theorem 10.5.
   (a) The projective plane of order 2.
   (b) The projective plane of order 3.
2. Let τ_1 and τ_2 be involutory collineations of a finite projective plane. Can $\tau_1\tau_2$ be fixed-point-free? Justify your answer.
3. Let σ be an automorphism of a projective plane $(\mathfrak{P},\mathfrak{L})$. Show that σ fixes all lines through some point P, iff σ fixes all points on some line ℓ. Thus σ has a centre iff σ has an axis.

   Note: In the finite case this follows from our classification of fixed substructures of collineations, which relies on the fact that σ has equally many fixed points and fixed lines (Corollary 10.3). However one verifies the required statement directly without any finiteness assumption. Suppose σ fixes every point of ℓ. We may assume that every fixed point of σ lies on ℓ (why?). Let R be a point not on ℓ, and let $P = (RR^\sigma) \cap \ell$. Then R lies on the fixed line RP; and so every point lies on at least one fixed line. Let S be a point not on $\ell \cup RP$ and consider how a fixed line through S may intersect RP.
4. How many polarities does the projective plane of order two have? Explain.
5. The Hermitian unital of order 2 consists of all absolute points and nonabsolute lines with respect to a unitary polarity of $\mathbb{P}^2(\mathbb{F}_4)$ as described above. List all points and lines of this unital, and draw it (showing all incidences). Where have you seen a design like this before in this course? and should you expect a similar connection between unitals and other designs we have studied, for larger values of q? Explain.

11. Classical Theorems

We present here several results that may be called *classical*, not only because of when these results were first proved; but because they describe properties of the classical projective planes $\mathbb{P}^2(F)$ not shared by more general projective planes.

The group $\Gamma L_3(F)$ of all semilinear transformations $F^3 \to F^3$ acts on the projective plane $\mathbb{P}^2(F)$. (See Appendix A2 for the notation and results of group theory used here.) This is because every semilinear transformation maps subspaces of F^3 to subspaces of the same dimension, preserving the incidence relation of inclusion. The normal subgroup $Z \trianglelefteq \Gamma L_3(F)$ consisting of all scalar transformations $F^3 \to F^3$, $v \mapsto \lambda v$ for $0 \neq \lambda \in F$, acts trivially on $\mathbb{P}^2(F)$ (i.e. it fixes every point and line) and so the quotient group

$$P\Gamma L_3(F) = \Gamma L_3(F)/Z$$

acts on $\mathbb{P}^2(F)$. The subgroup $PGL_3(F)$ consists of linear transformations $F^3 \to F^3$, but where two linear transformations are identified if one is a scalar multiple of the other (in which case their action on $\mathbb{P}^2(F)$ is the same).

An **ordered quadrangle** is an ordered 4-tuple of points (A, B, C, D), no three of which are collinear.

11.1 Fundamental Theorem of Projective Plane Geometry. The full collineation group of $\mathbb{P}^2(F)$ is $P\Gamma L_3(F)$. Its normal subgroup $PGL_3(F)$ is regular (i.e. sharply transitive) on ordered quadrangles.

Proof. We prove only the second statement. Recall that the points and lines of $\mathbb{P}^2(F)$ are the subspaces of F^3 of dimension 1 and 2 respectively. Let (A_1, A_2, A_3, A_4) be an ordered quadrangle, so that $A_i = \langle v_i \rangle$ for $i = 1, 2, 3, 4$; and no three of $v_1, v_2, v_3, v_4 \in V$ are linearly dependent. There exists a unique linear transformation $T : F^3 \to F^3$ such that $T(v_i) = e_i$ for $i = 1, 2, 3$ where we denote the standard basis of F^3 by

$$e_1 = (1, 0, 0); \quad e_2 = (0, 1, 0); \quad e_3 = (0, 0, 1).$$

Since $T(v_4)$ is not a linear combination of any two of $T(v_i) = e_i$ $(i = 1, 2, 3)$, we must have $T(v_4) = (a, b, c)$ where $abc \neq 0$. Define an invertible linear transformation $D : F^3 \to F^3$ by $(x, y, z) \mapsto (a^{-1}x, b^{-1}y, c^{-1}z)$. Then the invertible linear transformation $S \circ T : F^3 \to F^3$ maps

$$\begin{aligned} A_1 &\mapsto \langle (a^{-1}, 0, 0) \rangle = \langle (1, 0, 0) \rangle; \\ A_2 &\mapsto \langle (0, b^{-1}, 0) \rangle = \langle (0, 1, 0) \rangle; \\ A_3 &\mapsto \langle (0, 0, c^{-1}) \rangle = \langle (0, 0, 1) \rangle; \\ A_4 &\mapsto \langle (1, 1, 1) \rangle. \end{aligned}$$

Thus $PGL_3(F)$ is transitive on ordered quadrangles. Now suppose a linear transformation $R : F^3 \to F^3$ fixes the ordered quadrangle $(\langle e_1\rangle, \langle e_2\rangle, \langle e_3\rangle, \langle(1,1,1)\rangle)$. Since R fixes $\langle e_i\rangle$ for $i = 1, 2, 3$, the matrix of R is diagonal:

$$R = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

for some nonzero $a, b, c \in F$. Since R also fixes $\langle(1,1,1)\rangle$, we have

$$\langle(1,1,1)\rangle = R\langle(1,1,1)\rangle = \langle(a,b,c)\rangle$$

so that $a = b = c$ and $R \in Z$, which represents the identity in $PGL_3(F)$. Thus $PGL_3(F)$ permutes ordered quadrangles regularly (sharply transitively). □

The Fundamental Theorem is useful in making simplifying assumptions in any computations involving coordinates, as in the following proof.

11.2 Pappus' Theorem. In Figure 8.1 the points R_0, R_1, R_2 are collinear.

Proof. By the Fundamental Theorem, we may assume that

$$P_0{=}\langle(1,0,0)\rangle, \quad P_1{=}\langle(0,1,0)\rangle, \quad Q_0{=}\langle(0,0,1)\rangle, \quad Q_1{=}\langle(1,1,1)\rangle.$$

We immediately obtain

$$\ell{=}\langle(0,0,1)^T\rangle, \quad m{=}\langle(1,-1,0)^T\rangle, \quad O{=}\langle(1,1,0)\rangle, \quad P_2{=}\langle(1,\alpha,0)\rangle, \quad Q_2{=}\langle(1,1,\beta)\rangle$$

for some $\alpha, \beta \in F \smallsetminus \{0,1\}$. Next we determine

$$\begin{gathered} \ell_{01}{=}\langle(0,1,-1)^T\rangle, \quad \ell_{10}{=}\langle(1,0,0)^T\rangle, \quad \ell_{02}{=}\langle(0,-\beta,1)^T\rangle, \\ \ell_{20}{=}\langle(-\alpha,0,1)^T\rangle, \quad \ell_{12}{=}\langle(-\beta,0,1)^T\rangle, \quad \ell_{21}{=}\langle(-\alpha,1,\alpha{-}1)^T\rangle \end{gathered}$$

and consequently

$$R_0{=}\langle(1,\ \alpha{+}(1{-}\alpha)\beta,\ \beta)\rangle, \quad R_1{=}\langle(1,\alpha,\alpha\beta)\rangle, \quad R_2{=}\langle(0,1,1)\rangle.$$

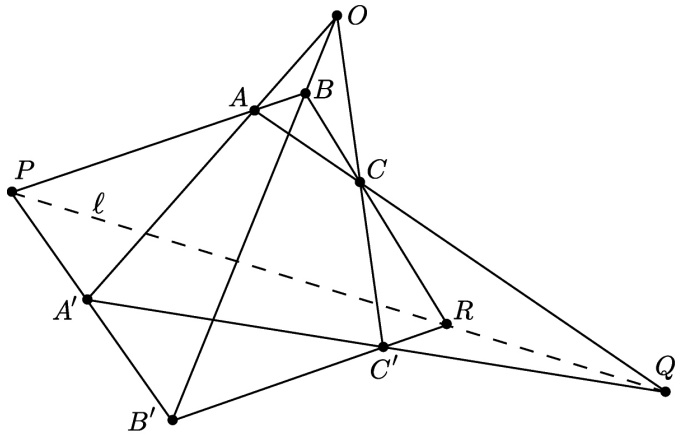
The fact that the latter three points are collinear follows from

$$\det \begin{bmatrix} 1 & \alpha{+}(1{-}\alpha)\beta & \beta \\ 1 & \alpha & \alpha\beta \\ 0 & 1 & 1 \end{bmatrix} = 0.$$

□

The following result may be proved using coordinates in a spirit similar to the previous proof. Instead our proof makes use of a plane collineation.

11.3 Desargues' Theorem. In a classical plane $\mathbb{P}^2(F)$ consider two triangles ABC and $A'B'C'$ sharing no sides or vertices, and suppose that the lines AA', BB' and CC' are concurrent at a point O. Let P, Q, R be the intersection points $AB \cap A'B'$, $AC \cap A'C'$, $BC \cap B'C'$ respectively, as in Figure 11.4. Then P, Q, R are collinear.



11.4 Figure
The Desargues Configuration

Proof. Let $\ell = PQ$. By comments following Proposition 10.6, there exists an (O, ℓ)-homology σ mapping $A \mapsto A'$. Since σ fixes both P and Q, while also fixing every line through O, it maps

$$B = PA \cap OB \mapsto PA' \cap OB = B';$$
$$C = QA \cap OC \mapsto QA' \cap OC = C';$$
$$BC \cap \ell \mapsto B'C' \cap \ell.$$

Since σ fixes every point of ℓ, this means that $BC \cap \ell = B'C' \cap \ell$, i.e. the three lines BC, $B'C'$, ℓ are concurrent; thus $R = BC \cap B'C'$ lies on ℓ as required. □

For each of the three preceding theorems, a converse is available. These converses, which are usually included in the statements of the theorems, state that the only planes with the given properties are the classical planes. A 'converse' of the Fundamental Theorem of Projective Plane Geometry, is the fact that every projective plane admitting an automorphism group which is transitive on ordered quadrangles, is necessarily classical. More is true: for example the Ostrom-Dembowski-Wagner Theorem states that a finite projective plane admitting an automorphism group which permutes the points 2-transitively, is necessarily classical.

Every projective plane which satisfies the conclusion of Desargues' Theorem, is isomorphic to $\mathbb{P}^2(K)$ for some skewfield K. (Recall that a skewfield is roughly a not-necessarily-commutative field; see Appendix A3 for a more complete description.) Every projective plane which satisfies the conclusion of Pappus' Theorem, is isomorphic to $\mathbb{P}^2(F)$ for some field F. Thus the condition of Pappus' Theorem is stronger than that of Desargues' Theorem: it implies commutativity of the underlying coordinate system. Planes of the form $\mathbb{P}^2(F)$ for some field F are called **Pappian**, while planes of the form $\mathbb{P}^2(K)$ for some skewfield K are called **Desarguesian**. Now by a *classical plane* we may intend either a Pappian plane or a more general Desarguesian plane; but for finite planes, these two concepts coincide since every *finite* skewfield is commutative (see Theorem A3.4). The alternative name **Galois plane** is also commonly applied to the finite classical planes $\mathbb{P}^2(\mathbb{F}_q)$ since they are coordinatized by the finite fields $\mathbb{F}_q$, also known as the **Galois fields** (and sometimes therefore denoted $GF(q)$).

For proofs of the converses of Desargues' Theorem and Pappus' Theorem, see e.g. [31]. These proofs rely on the coordinatization of a general projective plane, which is accomplished using a **planar ternary ring (PTR)**. Such a PTR consists of a set R with a ternary operation $T : R \times R \times R \to R$, rather than the two binary operations of addition and multiplication arising in traditional ring theory. In the classical case R is just a field, and $T(x, m, b) = xm + b$; but in the general case the ternary operation cannot be represented as a composite of two binary operations in such a convenient manner. The greater the degree of homogeneity, or symmetry, that is assumed for the projective plane, the nicer will be the algebraic properties of its PTR. For example translation planes, while not as symmetric as classical planes, are coordinatized by quasifields. A quasifield (defined in Section 3) is nicer than a general PTR, although not as nice as a field (at least it is typically neither commutative nor associative, and fails one of the distributive laws).

Exercises 11.

1. The Desargues configuration is the partial linear space shown in Figure 11.4 (with 10 points and 10 lines, including the line ℓ). Each point of the configuration lies on three lines of the configuration; and dually, each line passes through three points.
   (a) Show that the configuration is self-dual. (As a consequence, the dual of Desargues' Theorem can be phrased as a 'converse' of the original theorem; we omit the details.)
   (b) Identify the group of automorphisms of the configuration. (Partial credit will be given for providing generators and the order of the group. A complete answer will recognize the group as a known group.)

2. Prove that a classical projective plane $\mathbb{P}^2(F)$ has a subplane of order 2 iff the characteristic of F is 2.
   Hint. Use the Fundamental Theorem to choose nice coordinates for 4 of the 7 points of the subplane, without loss of generality.

3. Prove that the affine plane of order 3 embeds in $\mathbb{P}^2(\mathbb{F}_q)$, iff $q \not\equiv 2 \mod 3$.
   Hint. Again use the Fundamental Theorem to choose nice coordinates for 4 of the 9 points of the embedded affine plane, without loss of generality.

12. Conics and Ovals

A **conic** in a classical plane $\mathbb{P}^2(F)$ is the set $\mathcal{C}$ of points $\langle(x,y,z)\rangle$ satisfying a nonzero homogeneous quadratic polynomial condition of the form $Q(x,y,z)=0$ where

$$Q(X,Y,Z) = aX^2 + bY^2 + cZ^2 + dXY + eXZ + fYZ.$$

Note that this condition is well-defined for points $\langle(x,y,z)\rangle$ since if $0 \neq \lambda \in \mathbb{F}_q$ then $Q(\lambda x, \lambda y, \lambda z) = \lambda^2 Q(x,y,z)$. We say that $\mathcal{C}$ is **nondegenerate** if it does not contain an entire line. We show that for $F = \mathbb{F}_q$ a finite field, any two nondegenerate conics are equivalent by an invertible linear change of coordinates. Indeed we show that $\mathcal{C}$ can be transformed to the conic $y^2 = xz$ using a linear change of coordinates.

12.1 Theorem. Let $\mathcal{C}$ be a nondegenerate conic in a finite classical plane $\mathbb{P}^2(\mathbb{F}_q)$. Then $\mathcal{C}$ has $q+1$ points, of which no three are collinear. Moreover $\mathcal{C}$ is equivalent, by a linear change of coordinates, to the conic $y^2 = xz$.

Proof. By Theorem A1.7, $\mathcal{C}$ has at least one point P. By a linear change of coordinates, we may assume that $P = \langle(1,0,0)\rangle$; so without loss of generality $\mathcal{C}$ has the equation

$$bY^2 + cZ^2 + dXY + eXZ + fYZ = 0.$$

A typical line through P has the form

$$\ell_{\beta,\gamma} = \langle(0,\beta,\gamma)^T\rangle = \langle P, (0,\gamma,-\beta)\rangle \qquad \text{where } (\beta,\gamma) \neq (0,0).$$

Here we may take

$$(\beta,\gamma) \in \{(0,1)\} \cup \{(1,\gamma) : \gamma \in \mathbb{F}_q\}$$

in order to list the $q+1$ lines through P. We claim that

(12.2) exactly one of these lines (the tangent line at P) meets $\mathcal{C}$ at a unique point P; and the remaining q lines through P are secant lines, each meeting $\mathcal{C}$ at just one point other than P.

From (12.2) it will follow that $\mathcal{C}$ has exactly $q+1$ points; and that no line through P meets $\mathcal{C}$ in more than two points. But since P was an arbitrarily chosen point of $\mathcal{C}$, in fact no three points of $\mathcal{C}$ are collinear. To prove (12.2), consider an arbitrary point of $\ell_{\beta\gamma}$ other than P; this has the form $\langle(t,\gamma,-\beta)\rangle$ for some $t \in \mathbb{F}_q$. The condition for this point to lie on $\mathcal{C}$ is that

(12.3) $$b\gamma^2 - f\beta\gamma + c\beta^2 + (d\gamma - e\beta)t = 0.$$

If $(\beta,\gamma) \notin \langle (d,e)\rangle$ then the coefficient of t is nonzero so there is a unique solution for t. This gives a unique point of $\mathcal{C}$ on each line through P other than the line $\ell_{d,e}$. If $(\beta,\gamma) \in \langle (d,e)\rangle$ then the coefficient of t in (12.3) is zero and the equation becomes $be^2 + fde + cd^2 = 0$. If this equation holds then *all* points of the line $\ell_{d,e}$ lie on $\mathcal{C}$, contrary to the nondegeneracy of $\mathcal{C}$; therefore $be^2 + fde + cd^2 \neq 0$ and so $\mathcal{C}$ contains no points of $\ell_{d,e}$ other than P itself. This proves (12.2).

It follows from (12.2) that $\mathcal{C}$ has at least three noncollinear points P, Q, R. Let ℓ and m be the tangent lines at P and Q respectively, and let $S = \ell \cap m$; then no three of P, Q, R, S are collinear. By Theorem 11.1 we may choose coordinates so that $P = \langle (1,0,0)$; $S = \langle (0,1,0)\rangle$; $Q = \langle (0,0,1)\rangle$; $R = \langle (1,1,1)\rangle$. Now the equation of $\mathcal{C}$ takes the form

$$bY^2 + dXY + eXZ + fYZ = 0$$

where $b + d + e + f = 0$. The tangents to P and Q are given by $dY + eZ = 0$ and $eX + fY = 0$ respectively; since these tangents pass through $S = \langle (0,1,0)\rangle$ we must have $d = f = 0 \neq b$ and so the equation for $\mathcal{C}$ reduces further to $bY^2 + eXZ = 0$ where $e = -b \neq 0$. This gives the required form $Y^2 - XZ = 0$ for the equation of $\mathcal{C}$. □

A natural combinatorial generalization of a conic is provided by the following definition: A **k-arc** in a projective plane is a set $\mathcal{O}$ of k points of which no three are collinear. A line ℓ is called a **passant**, a **tangent** or a **secant** of $\mathcal{O}$ according as $|\ell \cap \mathcal{O}| = 0$, 1 or 2. An **oval** in a projective plane of order n is an $(n+1)$-arc. It is natural to ask what examples of ovals exist other than conics; and whether larger arcs exist than ovals.

12.4 Theorem. Let $\mathcal{O}$ be a k-arc in a projective plane of order n. If n is odd then $k \leq n+1$. If n is even then $k \leq n+2$.

Proof. Let $P \in \mathcal{O}$. Each of the $n+1$ lines through P has at most one point of $\mathcal{O}$ other than P itself, so that $k = |\mathcal{O}| \leq n+2$. Suppose there exists an $(n+2)$-arc; we must show that n is even. Let $P \in \mathcal{O}$. Since every line through P must be a secant line, $\mathcal{O}$ has no tangents. Let Q be a point *not* in $\mathcal{O}$; then every line through Q is either a passant line or a secant line. This means that $|\mathcal{O}| = n+2$ is even, so n is even. □

An $(n+2)$-arc in a plane of even order n is called a **hyperoval**.

Consider again the conic $XZ - Y^2 = 0$ in $\mathbb{P}^2(\mathbb{F}_q)$, whose points are given by

$$\mathcal{C} = \{\langle (1,t,t^2)\rangle : t \in \mathbb{F}_q\} \cup \{\langle (0,0,1)\rangle\}.$$

One checks that the tangents at these points are given by

$$\langle (t^2, -2t, 1)^T\rangle \quad \text{for } t \in \mathbb{F}_q; \qquad \langle (1,0,0)^T\rangle.$$

If q is odd then these tangents $\langle(\alpha,\beta,\gamma)^T\rangle$ satisfy $4\alpha\gamma-\beta^2=0$, i.e. they form a **dual conic** (a conic in the dual plane); in particular no three tangents are concurrent. However if q is even then the tangents appear more simply as

$$\langle(t^2,0,1)^T\rangle \quad \text{for } t\in\mathbb{F}_q; \qquad \langle(1,0,0)^T\rangle$$

and so all tangents pass through $N=\langle(0,1,0)\rangle$. Thus $\mathcal{C}\cup\{N\}$ is a hyperoval. This distinction in behaviour between conics in even and odd order generalizes to ovals in arbitrary finite planes for purely combinatorial reasons:

12.5 Theorem. Let $\mathcal{O}$ be an oval in a projective plane of order n. Then every point of $\mathcal{O}$ lies on a unique tangent; thus $\mathcal{O}$ has exactly $n+1$ tangents.

(a) If n is odd then no three tangents are concurrent.

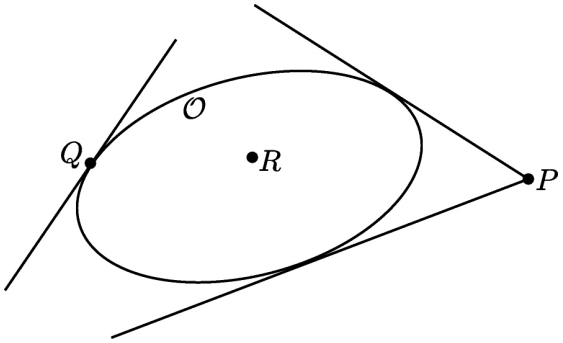
(b) If n is even then all $n+1$ tangents meet in a point N. Now $\mathcal{O}\cup\{N\}$ is a hyperoval, the unique hyperoval containing $\mathcal{O}$.

Proof. Let $P\in\mathcal{O}$. The n points $Q\in\mathcal{O}\smallsetminus\{P\}$ determine n distinct secants through P so the remaining line ℓ through P is a tangent.

Suppose n is odd. Every point $T\in\ell\smallsetminus\{P\}$ lies on an even number of tangents including ℓ since each tangent through T meets $\mathcal{O}$ in an odd number (i.e. one) of points, while the total number of points $|\mathcal{O}|=n+1$ is even. Since each of the n points $T\in\ell\smallsetminus\{P\}$ lies on at least one tangent other than ℓ, the tangents other than ℓ must meet ℓ in n distinct points. This proves (a).

Suppose n is even and let $Q\notin\mathcal{O}$. Since $|\mathcal{O}|=n+1$ is odd, Q must lie on an odd number of tangents. Since every point of $\mathcal{O}$ lies on a unique tangent, it follows that every point lies on at least one tangent. Let $\ell_0,\ell_1,\ldots,\ell_n$ be the tangents and let $N=\ell_0\cap\ell_1$. Let $\mathcal{S}$ be the set of points not covered by $\ell_0\cup\ell_1$, so that $|\mathcal{S}|=n(n-1)$. Each of the remaining tangents ℓ_j (for $j=2,3,\ldots,n$) covers either n or $n-1$ points of $\mathcal{S}$, according as ℓ_j does or does not pass through N. In order that the remaining $n-1$ tangents cover all $n(n-1)$ points of $\mathcal{S}$, we therefore require that they all pass through N. □

For an oval $\mathcal{O}$ in a plane of even order, the common point N of intersection of all tangents is called the **nucleus** or the **knot** of $\mathcal{O}$. For an oval $\mathcal{O}$ in a plane of odd order n, Theorem 12.5 shows that every point P is on 0, 1 or 2 tangents of $\mathcal{O}$; the point P is called **interior**, **absolute** or **exterior** with respect to an oval $\mathcal{O}$ accordingly. (An absolute point is simply a point of $\mathcal{O}$ itself.)



12.6 Figure
P is exterior;
Q is absolute;
R is interior

12.7 Theorem. Let $\mathcal{O}$ be an oval in a plane of *odd* order n. Then $\mathcal{O}$ has

$n{+}1$ absolute points, $\frac{1}{2}n(n{-}1)$ interior points, and $\frac{1}{2}n(n{+}1)$ exterior points;
$n{+}1$ tangents, $\frac{1}{2}n(n{-}1)$ passants, and $\frac{1}{2}n(n{+}1)$ secants.

Moreover

(a) Every absolute point lies on 1 tangent and n secants.
(b) Every interior point lies on $\frac{1}{2}\big(n{+}1\big)$ secants and $\frac{1}{2}\big(n{+}1\big)$ passants.
(c) Every exterior point lies on 2 tangents, $\frac{1}{2}\big(n{-}1\big)$ secants and $\frac{1}{2}\big(n{-}1\big)$ passants.
(d) Every tangent line contains 1 absolute point and n exterior points.
(e) Every passant line contains $\frac{1}{2}\big(n{+}1\big)$ exterior points and $\frac{1}{2}\big(n{+}1\big)$ interior points.
(f) Every secant line contains 2 absolute points, $\frac{1}{2}\big(n{-}1\big)$ exterior points and $\frac{1}{2}\big(n{-}1\big)$ interior points.

In the dual plane, the tangents of $\mathcal{O}$ become the points of a conic $\mathcal{O}'$. Duality interchanges the interior, absolute and exterior points of $\mathcal{O}$ (respectively, $\mathcal{O}'$) with the passant, tangent and secant lines of $\mathcal{O}'$ (respectively, $\mathcal{O}$).

Proof. The number of points and lines of each type follows by simple counting arguments; for example the $n{+}1$ tangents give $\binom{n+1}{2} = \frac{1}{2}n(n{+}1)$ distinct points of intersection, these being the exterior points. The number of interior points is therefore

$$(n^2{+}n{+}1) - (n{+}1) - \tfrac{1}{2}n(n{+}1) = \tfrac{1}{2}n(n{-}1)\,.$$

The final assertions concerning duality follow from Theorem 12.5(a). Conclusions (a), (b) and (c) follow from the definitions by straightforward counting. Statements (d), (e) and (f) follow by duality. □

Examples of ovals are known in most of the known finite projective planes. Segre's Theorem 12.14 shows that ovals in classical planes of odd order are necessarily conics. Not all finite projective planes have ovals; among the twenty-two known projective planes of order 16, four have no hyperovals and hence no ovals [54].

In planes of even order, every oval extends to a hyperoval; removing any point of a hyperoval yields an oval. So one naturally regards an oval as a hyperoval with one of its

points distinguished. Consider a conic $\mathcal{C}$ in a classical plane $\mathbb{P}^2(\mathbb{F}_q)$ where $q = 2^e$, and let N be its nucleus. The resulting hyperoval $\mathcal{O} = \mathcal{C} \cup \{N\}$ is called a **regular hyperoval**. Let G be the group of all collineations of $\mathbb{P}^2(\mathbb{F}_q)$ leaving invariant the hyperoval. For $q \leq 4$, the $q+2$ points of the hyperoval are permuted transitively by G and all ovals are equivalent under G. However for $q \geq 8$, the group G has two orbits on the points of the hyperoval, namely $\mathcal{C}$ and $\{N\}$. In this case two types of oval are available from $\mathcal{O}$, namely a conic $\mathcal{C}$ and a '**pointed conic**' of the form $\mathcal{O}\frown\{P\}$ where $P \in \mathcal{C}$. It is sufficient, and clearly easier, to focus our attention on hyperovals rather than ovals when studying planes of even order.

For $q \leq 8$, every hyperoval in $\mathbb{P}^2(F_q)$ is regular. Irregular hyperovals exist in $\mathbb{P}^2(\mathbb{F}_q)$ for even $q \geq 16$. The unique irregular hyperoval in $\mathbb{P}^2(\mathbb{F}_{16})$ is due to Lunelli and Sce [41]. We provide here a very brief introduction to hyperovals; for a more complete survey see, for example, [18].

Let $\mathcal{O}$ be a hyperoval in $\mathbb{P}^2(\mathbb{F}_q)$ where $q = 2^e \geq 4$. By an appropriate choice of coordinates we may assume that $\mathcal{O}$ contains the points $P = \langle(1,0,0)\rangle$ and $Q = \langle(0,1,0)\rangle$. None of the remaining q points of $\mathcal{O}$ may lie on the line $PQ = \langle(0,1,0)^T\rangle$, so they all have the form $\langle(s,t,1)\rangle$ with distinct s's (since no two of them are collinear with P) and distinct t's (since no two of them are collinear with Q). Thus

$$\mathcal{O} = \{\langle(f(t),t,1)\rangle \,:\, t \in \mathbb{F}_q\} \cup \{\langle(1,0,0)\rangle, \langle(0,1,0)\rangle\}$$

for some permutation $f : \mathbb{F}_q \to \mathbb{F}_q$. It is convenient to regard $f(X)$ as represented by a polynomial in $\mathbb{F}_q[X]$, which must therefore be a permutation polynomial; see Appendix A2. We regard PQ as the line at infinity, and the points of $\mathcal{O}$ outside this line as the graph of f in the affine plane $\mathbb{A}^2(\mathbb{F}_q)$. Since no three affine points of $\mathcal{O}$ are collinear, we have

$$\det \begin{bmatrix} f(r) & r & 1 \\ f(s) & s & 1 \\ f(t) & t & 1 \end{bmatrix} \neq 0$$

whenever $r, s, t \in \mathbb{F}_q$ are distinct; equivalently,

$$\frac{f(r)-f(t)}{r-t} \neq \frac{f(s)-f(t)}{s-t} \text{ whenever } r,s,t \in \mathbb{F}_q \text{ are distinct.} \tag{12.8}$$

Of course all $-$ signs here are just $+$ signs in characteristic two; but we write them as $-$ signs so that the quotients in (12.8) will be clearly recognized as slopes of secants in the graph of f. We have reduced the search for hyperovals, to the search for permutations satisfying the condition (12.8). Again it is convenient to regard $f(X)$ as a polynomial of degree less than q. Accordingly, any permutation polynomial satisfying (12.8) is called an ***o*-polynomial**. Often it is convenient to add the conditions that $f(0) = 0$ and $f(1) = 1$. These conditions we may assume, since with appropriate choice of coordinates we may also assume that $\mathcal{O}$ contains $\langle(0,0,1)\rangle$ and $\langle(1,1,1)\rangle$.

12.9 Example: Regular hyperovals. Let $q = 2^r$, $f(t) = t^2$. The resulting hyperoval consists of the conic $XZ - Y^2 = 0$ together with the nucleus $\langle(0,1,0)\rangle$. Note that f is simply the generator of $\mathrm{Aut}\,\mathbb{F}_q = \langle\sigma\rangle = \{1, \sigma, \sigma^2, \ldots, \sigma^{r-1}\}$, $\sigma : x \to x^2$. One checks directly that (12.8) reduces to the condition that $r + t \neq s + t$ whenever r, s, t are distinct. This condition is clearly satisfied, as well as the condition that f be bijective.

One may also let $f(t) = t^{1/2} = t^{2^{r-1}}$, i.e. $f = \sigma^{-1} = \sigma^{r-1}$. In this case $\mathcal{O}$ is clearly the conic $X^2 - YZ = 0$ together with its nucleus $\langle(1,0,0)\rangle$. Here the condition (12.8) becomes

$$\frac{1}{r^{1/2} + t^{1/2}} = \frac{r^{1/2} + t^{1/2}}{r + t} \neq \frac{s^{1/2} + t^{1/2}}{s + t} = \frac{1}{s^{1/2} + t^{1/2}}$$

whenever r, s, t are distinct. Raising both sides to the power -2 gives $r + t \neq s + t$ as before.

The latter example illustrates the more general observation that if f is an o-polynomial, then the inverse function f^{-1} is also an o-polynomial whose hyperoval is equivalent to that defined by f, by an interchange of the x and y coordinates.

12.10 Example: Segre hyperovals. Let $q{=}2^r$, $f(t){=}t^{2^k}$ for some $k \in \{0, 1, 2, \ldots, r{-}1\}$ so that $f = \sigma^k$ is a field automorphism. Thus f is automatically bijective. Since f is a field automorphism, (12.8) becomes that

$$\frac{(r-t)^{2^k}}{r-t} \neq \frac{(s-t)^{2^k}}{s-t}$$

whenever r, s, t are distinct; equivalently, the map $x \mapsto x^{2^k-1}$ is bijective. This means that $gcd(2^k - 1, 2^r - 1) = 2^{gcd(k,r)} - 1 = 1$, i.e. $gcd(k, r) = 1$. The resulting hyperovals are irregular if $k \notin \{1, r{-}1\}$. The smallest example arises for $r = 5$ and $k = 2$, which gives an irregular Segre hyperoval

$$\{\langle(t^4, t, 1)\rangle \,:\, t \in \mathbb{F}_{32}\} \cup \{\langle(1,0,0)\rangle, \langle(0,1,0)\rangle\}$$

in $\mathbb{P}^2(\mathbb{F}_{32})$. By the preceding remarks, the hyperoval arising from the o-polynomial $t^8 = f^{-1}(t)$ is equivalent to the one arising from f.

12.11 Example: The Lunelli-Sce Hyperoval. Let $q = 16$ and write $\mathbb{F}_{16} = \mathbb{F}_2[\omega]$ where $\omega^4 = \omega + 1$ (see Appendix A2.3). The unique irregular hyperoval in $\mathbb{P}^2(\mathbb{F}_{16})$ may be defined using the o-polynomial

$$f(t) = t^{12} + t^{10} + \omega^{11}t^8 + t^6 + \omega^2 t^4 + \omega^9 t^2.$$

Don't even think about trying to verify by hand that this is an o-polynomial! Fortunately this is an easy job using appropriate computer software. The remarkable thing is that Lunelli and Sce discovered this hyperoval in 1958 by computer!

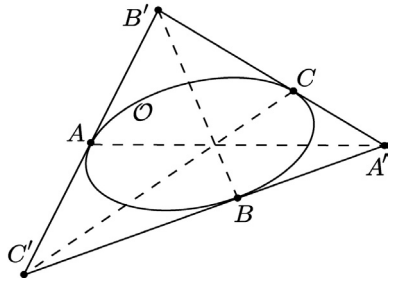
The subject of hyperovals (and connections to flocks and q-clans) is too big a subject to cover here! But I can't resist giving a couple further examples.

12.12 Example: Trinomial Hyperovals. We describe two infinite families of hyperovals for $q = 2^{2e+1}$, the **Payne hyperovals** and the **Cherowitzo hyperovals**. Both are described using o-polynomials with only three terms. Here $\mathbb{F}_q^\times$ is cyclic of order $q-1 \equiv 1 \mod 6$, so the map $x \mapsto x^6$ is bijective, which means that every $x \in \mathbb{F}_q$ has a unique sixth root $x^{1/6}$. This is a monomial in x if the exponent is interpreted mod $(q-1)$. Define $f(x) = x^{1/6} + x^{3/6} + x^{5/6}$. Payne [52] showed that this is an o-polynomial.

Also, $\mathrm{Aut}(\mathbb{F}_q)$ is cyclic of odd order $2e+1$. Therefore every automorphism of $\mathbb{F}_q$ has a unique square root. Let $\sigma \in \mathrm{Aut}(\mathbb{F}_q)$ such that σ^2 is the usual generator $x \mapsto x^2$ of $\mathrm{Aut}(\mathbb{F}_q)$. Then $f(x) = x^\sigma + x^{\sigma+2} + x^{3\sigma+4}$ is an o-polynomial [17]. Cherowitzo conjectured this result in 1985 based on computational evidence for small values of e. His proof of this result, a decade later, was quite a tour de force.

Finally we show that ovals in classical planes of odd order are conics.

12.13 Lemma. Let $\mathcal{O}$ be an oval in $\mathbb{P}^2(\mathbb{F}_q)$ where q is odd. Let A, B, C be three distinct points of $\mathcal{O}$ as shown, and let A', B', C' be the vertices of the triangle formed by the tangents at A, B, C, with A' opposite A, etc. Then the three lines AA', BB', CC' are concurrent.



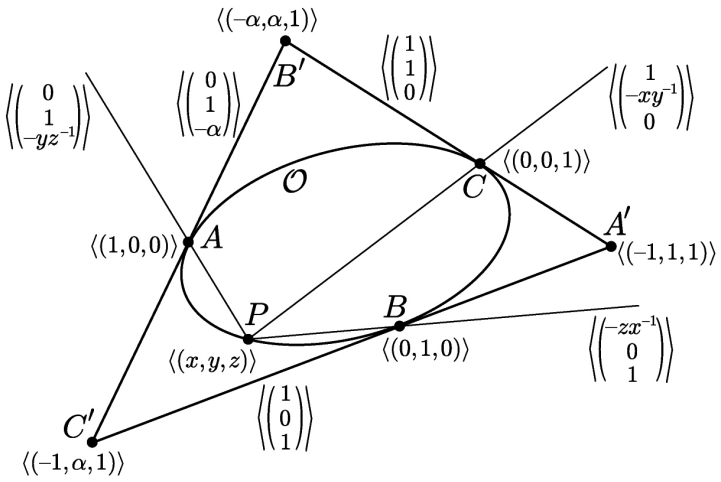
Proof. We may choose coordinates such that

$$A = \langle(1,0,0)\rangle, \quad B = \langle(0,1,0)\rangle, \quad C = \langle(0,0,1)\rangle, \quad A' = \langle(-1,1,1)\rangle.$$

The tangents to $\mathcal{O}$ at A, B, C are

$$\langle(0,1,-\alpha)^T\rangle, \quad \langle(1,0,1)^T\rangle, \quad \langle(1,1,0)^T\rangle$$

respectively, for some $\alpha \neq 0$; this means that $B' = \langle(-\alpha,\alpha,1)\rangle$, $C' = \langle(-1,\alpha,1)\rangle$.



Consider an arbitrary point $P \in \mathcal{O} \smallsetminus \{A, B, C\}$, so that P has coordinates $\langle (x, y, z) \rangle$ where $xyz \neq 0$. The secants AP, BP, CP have coordinates

$$\langle (0, 1, -yz^{-1})^T \rangle, \quad \langle (-zx^{-1}, 0, 1)^T \rangle, \quad \langle (1, -xy^{-1}, 0)^T \rangle$$

respectively, where the product of the three nontrivial entries satisfies

$$(-yz^{-1})(-zx^{-1})(-xy^{-1}) = -1.$$

Now consider a triple of lines of the form

$$\ell_{A,a} = \langle (0, 1, a)^T \rangle, \quad \ell_{B,b} = \langle (b, 0, 1)^T \rangle, \quad \ell_{C,c} = \langle (1, c, 0)^T \rangle$$

such that $abc \neq 0$. This gives an arbitrary triple of lines distinct from the sides of the triangle ABC, but passing through A, B, C respectively. These lines must correspond (in some order) to the secants and tangents at A, B, C considered above. Therefore

$$\Big(\prod_{a \in \mathbb{F}_q^\times} a \Big) \Big(\prod_{b \in \mathbb{F}_q^\times} b \Big) \Big(\prod_{c \in \mathbb{F}_q^\times} c \Big) = \underbrace{(-\alpha)(1)(1)}_{\text{tangents}} \underbrace{(-1)^{q-2}}_{\text{secants}}.$$

Here the first three factors arise from the fact that the tangents at A, B, C are $\ell_{A,-\alpha}$, $\ell_{B,1}$, $\ell_{C,1}$ respectively. The latter expression $(-1)^{q-2}$ arises from the $q-2$ triples of secants of the form AP, BP, CP, each of which contributes a factor -1 to the product as we have seen. By Exercise #2 each of the three products on the left is -1, so we obtain $\alpha = -1$. Finally we obtain

$$AA' = \langle (0, 1, -1)^T \rangle, \quad BB' = \langle (-1, 0, 1)^T \rangle, \quad CC' = \langle (1, -1, 0)^T \rangle$$

and these three lines all meet at a point $\langle (1, 1, 1) \rangle$ as required. □

12.14 Segre's Theorem. If $\mathcal{O}$ is an oval in a finite classical plane $\mathbb{P}^2(\mathbb{F}_q)$ where q is odd, then $\mathcal{O}$ is a conic.

Proof. We use the same notation as in the proof of Lemma 12.13. Denote the tangent to $\mathcal{O}$ at $P = \langle (x, y, z) \rangle$ by $\langle (a, b, c) \rangle$ and consider the triangle formed by the tangents at B, C, P. This triangle has vertices

$$\langle (c, -c, b-a) \rangle, \quad \langle (b, c-a, -b) \rangle, \quad \langle (-1, 1, 1) \rangle$$

opposite B, C, P respectively, as shown. The secants joining B, C, P to the opposite vertices of the triangle are

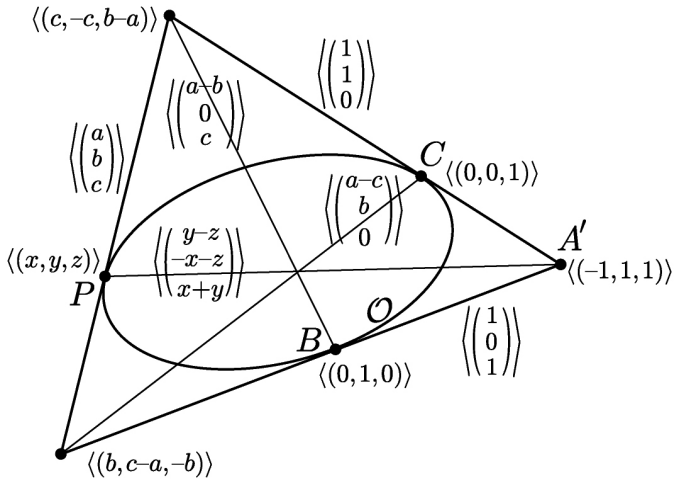
$$\langle (a-b, 0, c)^T \rangle, \quad \langle (a-c, b, 0)^T \rangle, \quad \langle (y-z, -x-z, x+y)^T \rangle$$

respectively. By Lemma 12.13 these three secants are concurrent, so that

$$0 = \det \begin{bmatrix} a-b & a-c & y-z \\ 0 & b & -x-z \\ c & 0 & x+y \end{bmatrix} = (a - b - c)(bx + by - cx - cz).$$

Since $A' = \langle (-1, 1, 1) \rangle$ does not lie on the secant $\langle (a-c, b, 0)^T \rangle$ through C, we have

$$b(x+y) = c(x+z).$$



Applying the same argument to the triples ACP and ABP in place of BCP, and exploiting the symmetry in the three coordinates, we conclude that $(y{+}z,\ x{+}z,\ x{+}y) = \lambda(a,b,c)$ for some $\lambda \in \mathbb{F}_q^\times$; thus

$$2(xy + xz + yz) = \lambda(ax + by + cz) = 0.$$

Thus every point $P \in \mathcal{O} \smallsetminus \{A, B, C\}$, as well as the three points A, B, C, lie on the conic $\mathcal{C}$ defined by $xy{+}xz{+}yz = 0$. It follows that $\mathcal{O} = \mathcal{C}$. □

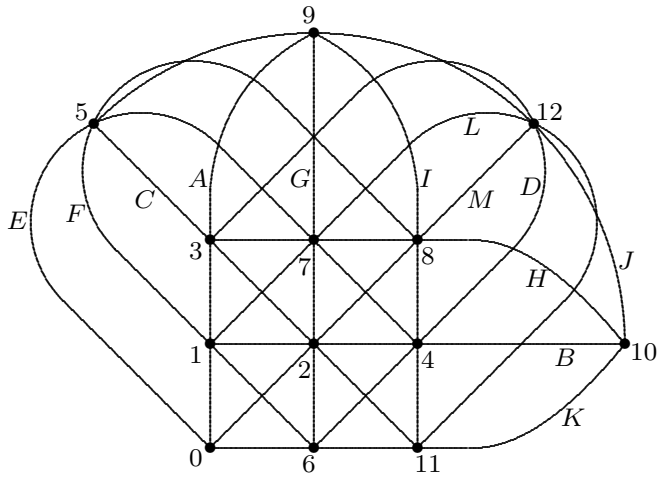
Exercises 12.

1. How many ovals and hyperovals does a projective plane of order 2 have? Show that all ovals in this plane are conics.

2. Show that the product of all nonzero elements in any finite field is -1. (This fact is used in the proof of Lemma 12.13.)

   Hint: Most terms in this product occur in pairs a, a^{-1}.

3. Consider the oval $\mathcal{O} = \{0, 1, 2, 6\}$ in the projective plane of order 3 with point set $\mathfrak{P} = \{0, 1, 2, \ldots, 12\}$ and line set $\mathfrak{L} = \{A, B, \ldots, M\}$ as labelled below.



(a) List the interior points; absolute points; exterior points; passant lines; tangent lines; and secant lines for $\mathcal{O}$.

(b) Complete the following sentences correctly, using your sketch of the plane:

Every interior point lies on _____ passants, _____ tangents, and _____ secants;
every absolute point lies on _____ passants, _____ tangents, and _____ secants;
every exterior point lies on _____ passants, _____ tangents, and _____ secants.

Every passant contains _____ interior points, _____ absolute points, and _____ exterior points;
every tangent contains _____ interior points, _____ absolute points, and _____ exterior points;
every secant contains _____ interior points, _____ absolute points, and _____ exterior points.

4. Let $\ell_1, \ell_2, \ldots, \ell_k$ be $k \geq 1$ distinct lines in a projective plane $(\mathfrak{P}, \mathfrak{L})$ of order n. Suppose that for every point $P \in \mathfrak{P}$, the number of $i \in \{1, 2, \ldots, k\}$ such that $P \in \ell_i$ is even.

(a) Show that $k \geq n+2$.

(b) Can $k = n+2$? Does your answer depend on n or the choice of plane $(\mathfrak{P}, \mathfrak{L})$? Explain.

Hint. (a) Consider the intersections of $\ell_2, \ldots, \ell_k$ with ℓ_1. (b) Think: hyperovals.

13. Codes of Planes

Later in this section, we will require basic terminology and results of coding theory and invariant theory, as briefly summarized in Appendices A5 and A6. We begin, however, by introducing the Smith normal form of an arbitrary integer matrix.

13.1 Theorem. Let A be an $m \times n$ matrix with integer entries. Then there is a unique $m \times n$ matrix $D = \big[d_{ij}\big]$ (the **Smith normal form** of A) satisfying $MAN = D$ where M and N are integer matrices of size $m \times m$ and $n \times n$ respectively, each having determinant ± 1, such that the entries of D are non-negative integers satisfying $d_{ij} = 0$ for $i \neq j$ (i.e. D is 'diagonal') and $d_{11} \,\big|\, d_{22} \,\big|\, \cdots \,\big|\, d_{\mu\mu}$ where $\mu = \min\{m, n\}$.

Note that if any of the diagonal entries d_{ii} is zero, then all remaining diagonal entries d_{jj} for $j \geq i$ must also vanish. The diagonal entries d_{ii}, which uniquely determine the Smith normal form of A, are called the **elementary divisors** of A. The *uniqueness* of the Smith normal form is shown in Exercise #5. The existence follows from Exercise #6, which in fact gives an algorithm for computing the Smith normal form. This algorithm starts with the observation that d_{11} equals the greatest common divisor of all entries of A. As a first example, the randomly generated matrix

$$A_0 = \begin{bmatrix} -20 & -16 & 22 & -12 & 44 \\ -33 & -43 & 40 & -8 & 56 \\ -14 & -40 & 22 & 12 & 2 \\ -25 & -13 & 26 & -20 & 62 \end{bmatrix}$$

may be factored as

$$A_0 = \begin{bmatrix} 6 & 7 & 11 & -8 \\ 11 & 11 & 20 & -15 \\ 6 & 4 & 11 & -9 \\ 7 & 9 & 13 & -9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 & 0 & 0 & 2 \\ -1 & 4 & 0 & -4 & 7 \\ 0 & -3 & 1 & 2 & -3 \\ 1 & 0 & 0 & 1 & -3 \\ -1 & 0 & 0 & 0 & 2 \end{bmatrix}$$

where the square matrices shown have determinant -1 and 1 respectively; thus the elementary divisors of A_0 are $1, 2, 2, 0$, which we abbreviate as $1^1 2^2 0^1$.

Once again let A be an arbitrary $m \times n$ matrix with integer entries. For any prime p, the **p-rank** of A is the rank of A over any field of characteristic p. (The rank of the matrix A over a field F, depends only on the matrix A and the characteristic of F.) For example it is clear from the Smith normal form given above, that the matrix A_0 has p-rank equal to 3 (the same as the rank over fields of characteristic zero) for $p \neq 2$, whereas its 2-rank equals 1. More generally the p-rank of A is the number of elementary divisors of A which are not divisible by p.

If A is an incidence matrix of an incidence structure with m points and n blocks, then we may refer to the p-rank of A as the p-rank of the incidence structure itself. Although there are in general exponentially many choices of incidence matrix for a given structure, arising from permutations of the rows and columns of A, these permutations will not affect the p-rank of A, and so the p-rank of the incidence structure is well-defined. The p-rank serves as a useful isomorphism invariant, since it is easily computed by Gaussian elimination. This is because the only matrix entries which arise during intermediate computations are $0, 1, 2, \ldots, p-1$. By comparison, computing the rank of a matrix over $\mathbb{Q}$ is typically very expensive, since there is no bound on the complexity of the rational entries which arise during intermediate computations. Similarly, the algorithm suggested in Exercise #6 for computing Smith normal forms typically leads to excessively large matrix entries during intermediate stages of the computation, even if the resulting elementary divisors turn out to be rather small. Fortunately there are other algorithms which help to overcome this obstacle when computing Smith normal forms. The elementary divisors of the incidence structure are also an isomorphism invariant, which in general give stronger information than the p-ranks themselves, since given the elementary divisors, one can easily read off the p-ranks for all primes p.

For example two of the four projective planes of order 9 have p-ranks as listed for small primes p:

plane	**2-rank**	**3-rank**	**5-rank**	**7-rank**	**elementary divisors**
$PG_2(9)$	90	37	90	91	$1^{37}3^{18}9^{35}90^1$
Hall plane of order 9	90	41	90	91	$1^{41}3^{10}9^{39}90^1$

The Smith normal form (or the 3-rank, which is much easier to compute) suffices to prove that these two planes are non-isomorphic. The p-ranks for primes p not dividing the order of the plane, however, carry no information useful in distinguishing the plane; they depend only on the order of the plane. See Exercise #4 for an explanation of this phenomenon. Unfortunately it is not always possible to distinguish non-isomorphic objects using p-ranks alone; for example the Hughes plane of order 9 has the same p-ranks as the Hall plane of order 9. A good rule of thumb is that more classical (or symmetric, or nicer) objects tend to have lower p-ranks for the relevant primes, as well as larger automorphism groups.

Let A be an incidence matrix of a projective plane of order two. Any such incidence matrix will do; we will take

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

We check that the elementary divisors of A (i.e. of the plane of order 2) are $1^4 2^2 6^1$, and in particular its 2-rank is 4. Let $\mathcal{C} \leq \mathbb{F}_2^7$ be the row space of A over $\mathbb{F}_2$, so that $\dim \mathcal{C} = 4$ and $|\mathcal{C}| = 2^4 = 16$. The sixteen vectors of $\mathcal{C}$ are listed in the table in Appendix A5.8; that is, $\mathcal{C}$ is the $[7, 4, 3]$ binary Hamming code as presented there. Its codewords of weight three and four are simply the rows of A and their complements (obtained by switching $0 \leftrightarrow 1$); together with the zero vector and the unique vector of weight 7, these account for all 16 codewords. Different choices of incidence matrix for the projective plane of order 2 give Hamming codes with the same parameters, differing from our code $\mathcal{C}$ only by a permutation of the seven coordinates. Conversely, every binary $[7, 4, 3]$-code has exactly seven codewords of weight three, and these form the rows of an incidence matrix for the projective plane of order 2. This connection between the smallest projective plane and the (arguably) smallest interesting code, a perfect code. But there is more...

Let $\widehat{\mathcal{C}} \leq \mathbb{F}_2^8$ be the linear code obtained from $\mathcal{C}$ by adding one extra coordinate as a parity check. Comparing with the list in Appendix A5.8 we see that

$$\widehat{\mathcal{C}} = \{00000000,\ 00011110,\ 00101101,\ 00110011,\ \ldots,\ 11111111\}.$$

Note that $\widehat{\mathcal{C}}$ has 14 words of weight 4, together with the zero vector and the vector 11111111 of weight 8; thus the weight enumerator of of $\widehat{\mathcal{C}}$ is

$$A_{\widehat{\mathcal{C}}}(x, y) = x^8 + 14x^4y^4 + y^8.$$

This code $\widehat{\mathcal{C}}$ is the $[8,4,4]$ **extended binary Hamming code**. A generator matrix for $\widehat{\mathcal{C}}$ is given by the 7×8 augmented matrix

$$G = \begin{bmatrix} A & \mathbf{1}^T \end{bmatrix}$$

where $\mathbf{1} = (1,1,1,1,1,1,1)$. Since

$$GG^T \;=\; AA^T \,+\, \mathbf{1}^T\mathbf{1} \;=\; 7I + J + J \;=\; 0$$

(the 7×7 matrix of zeroes over $\mathbb{F}_2$) it follows that $\widehat{\mathcal{C}} \subseteq \widehat{\mathcal{C}}^\perp$; but then comparing dimensions, that $\widehat{\mathcal{C}}$ is *self-dual;* that is, $\widehat{\mathcal{C}}^\perp = \widehat{\mathcal{C}}$. In particular from the MacWilliams relations A5.4 we have

$$A_{\widehat{\mathcal{C}}}(x,y) \;=\; A_{\widehat{\mathcal{C}}^\perp}(x,y) \;=\; \tfrac{1}{16} A_{\widehat{\mathcal{C}}}(x+y,\, x-y)\,,$$

i.e.

$$x^8 + 14x^4y^4 + y^8 = \tfrac{1}{16}\big[(x+y)^8 + 14(x+y)^4(x-y)^4 + (x-y)^8\big],$$

which can be easily verified by routine expansion.

Several generalizations of this construction of the $[7,4,3]$ binary Hamming code are available:

- If $q = p^r$ then the code (over any field of characteristic p) of the projective plane $\mathbb{P}^2(\mathbb{F}_q)$ has dimension $\binom{p+1}{2}^r + 1$. This is a special case of Theorem 22.1. However, the weight enumerators of these codes are not known in general. Indeed, no weight enumerator has currently been computed for any plane of order exceeding 8. The problem of computing weight enumerators of *general* codes, is recognized as exceedingly hard: no polynomial-time algorithm for this is known, or is likely to exist.
- The $[7,4,3]$ binary Hamming code is the first member of an infinite family of perfect 1-error correcting binary Hamming codes which may be constructed from higher-dimensional projective spaces over $\mathbb{F}_2$; see Exercise #19.5.
- If a prime p divides the order n of a projective plane, then the extended code of the plane over any field of characteristic p is self-orthogonal (i.e. $\widehat{\mathcal{C}} \subseteq \widehat{\mathcal{C}}^\perp$) with respect to an appropriately chosen symmetric bilinear form. Now suppose that p **sharply divides** n, i.e. $p \,\big|\, n$ but $p^2 \nmid n$; this condition is denoted $p \,\big\|\, n$. In this case we will show that equality holds, i.e. $\widehat{\mathcal{C}} = \widehat{\mathcal{C}}^\perp$, so that $\widehat{\mathcal{C}}$ is self-dual with respect to B. These facts we proceed to explain below.

Let A be an incidence matrix for a projective plane of order n, and let p be any prime dividing n. The extended code $\widehat{\mathcal{C}}$ of the plane, over the field $\mathbb{F}_p$, has generator matrix

$$G = \begin{bmatrix} A & -\mathbf{1}^T \end{bmatrix}$$

of size $N \times (N{+}1)$ where $N = n^2{+}n{+}1$ and $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{F}_p^N$. Using the symmetric bilinear form on $\mathbb{F}_p^{N+1}$ defined by

$$B(x,y) = xDy^T = x_1y_1 + x_2y_2 + \cdots + x_Ny_N - x_{N+1}y_{N+1}, \quad D = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix}$$

Any two codewords have the form $xG, yG \in \widehat{\mathcal{C}}$ for some $x, y \in \mathbb{F}_p^N$, so that

$$B(xG, yG) = xGDG^Ty^T = x(AA^T - \mathbf{1}^T\mathbf{1})y^T = x(nI + J - J)y^T = 0$$

since $p \,\big|\, n$. This shows that $\widehat{\mathcal{C}}$ is self-orthogonal with respect to B. Equivalently,

(13.2) $\widehat{\mathcal{C}}D \subseteq \widehat{\mathcal{C}}^\perp$ where $\perp$ is the 'perp' with respect to the standard dot product on $\mathbb{F}_p^{N+1}$. In particular, $\dim \widehat{\mathcal{C}} \le \frac{1}{2}\big(N{+}1\big) = \frac{1}{2}\big(n^2{+}n{+}2\big)$.

We will show that equality in (13.2), i.e. $\widehat{\mathcal{C}}$ is self-dual, whenever $p \,\big\|\, n$:

13.3 Theorem. If a prime p sharply divides n, then every projective plane of order n has p-rank equal to $\frac{1}{2}\big(n^2{+}n{+}2\big)$. Moreover the extended code $\widehat{\mathcal{C}}$ is self-dual (with respect to the form B as indicated above).

Proof. It remains only to be shown that $\dim \widehat{\mathcal{C}}^\perp = \frac{1}{2}\big(N+1\big)$. Since $\mathbf{1}^T$ lies in the column space of A, the matrix G has the same p-rank as A, i.e. $\dim \widehat{\mathcal{C}} = \dim \mathcal{C}$. Since

$$(\det A)^2 = \det(AA^T) = \det(nI + J) = (n+1)^2 n^{n(n+1)}$$

by Exercise #2, we have

$$|\det A| = (n{+}1)n^{n(n+1)/2} = d_1 d_2 \cdots d_N$$

where $d_1, d_2, \ldots, d_N$ are the elementary divisors of A. Since $p \,\big\|\, n$, at most $\frac{1}{2}\big(n^2{+}n\big)$ of the d_i's are divisible by p, so that the p-rank of A is at least $\frac{1}{2}\big(n^2{+}n{+}2\big) = \frac{1}{2}\big(N{+}1\big)$. This is the reverse of the inequality (13.2), so that in fact equality must hold. □

13.4 Corollary. Let p be a prime sharply dividing n, and suppose $\widehat{\mathcal{C}}$ is the extended code of length $N{+}1 = n^2{+}n{+}2$ (over $\mathbb{F}_p$) of a projective plane of order n. Then the weight enumerator $A_{\widehat{\mathcal{C}}}(z)$ satisfies

$$A_{\widehat{\mathcal{C}}}(x,y) \;=\; \frac{1}{p^{(N+1)/2}} A_{\widehat{\mathcal{C}}}\big(x + (p{-}1)y,\, x - y\big) = A_{\widehat{\mathcal{C}}}\Big(\frac{x + (p{-}1)y}{\sqrt{p}},\, \frac{x-y}{\sqrt{p}}\Big).$$

Proof. By Theorem 13.3, equality holds in (13.2), i.e. $\widehat{\mathcal{C}}^\perp = \widehat{\mathcal{C}}D$. But D simply multiplies the last coordinate of every codeword by -1, and this does not change the weight of any word. Thus $\widehat{\mathcal{C}}^\perp$ has the same weight enumerator as $\widehat{\mathcal{C}}$. The result follows from the MacWilliams relations A5.4. □

Henceforth we assume for simplicity that $p = 2$; in this case the result of Corollary 13.4 reads

$$A_{\widehat{\mathcal{C}}}(x,y) = A_{\widehat{\mathcal{C}}}\Big(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}\Big). \tag{13.5}$$

In this case we also have

$$A_{\widehat{\mathcal{C}}}(x,y) = A_{\widehat{\mathcal{C}}}(y,x), \tag{13.6}$$

which simply expresses the fact that $A_{n-d} = A_d$: the map $w \leftrightarrow \mathbf{1} + w$ gives a bijection between codewords of weight d and codewords of weight $n-d$. Thus the homogeneous polynomial $A_{\widehat{\mathcal{C}}}(x,y) \in \mathbb{C}[x,y]$ of degree $N+1$ is invariant under the group

$$G = \big\langle \tfrac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right],\ \left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]\big\rangle.$$

We check that this is a group of order 16:

$$\begin{aligned} G = \big\{ &\pm\left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right],\ \pm\left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right],\ \pm\left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right],\ \pm\left[\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right],\\ &\pm\tfrac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right],\ \pm\tfrac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & -1\\ 1 & 1\end{smallmatrix}\right],\ \pm\tfrac{1}{\sqrt{2}}\left[\begin{smallmatrix}-1 & 1\\ 1 & 1\end{smallmatrix}\right],\ \pm\tfrac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ -1 & 1\end{smallmatrix}\right]\big\}. \end{aligned}$$

By Molien's Theorem A6.4 the ring of invariants $\mathbb{C}[x,y]^G$ has Hilbert series

$$\begin{aligned} H_{\mathbb{C}[x,y]^G}(t) &= \frac{1}{16}\Big[\frac{1}{(1-t)^2} + \frac{1}{(1+t)^2} + \frac{8}{1-t^2} + \frac{2}{1+t^2} + \frac{2}{1-\sqrt{2}t+t^2} + \frac{2}{1+\sqrt{2}t+t^2}\Big]\\ &= \frac{1}{(1-t^2)(1-t^8)}\,. \end{aligned}$$

By Theorem A6.5 we will be done if we can find two algebraically independent invariants η_1 and η_2 of degree 2 and 8 respectively. Since all matrices in G are real orthogonal, it is clear (cf. Example A6.2) that x^2+y^2 is invariant; let us take $\eta_1(x,y) = x^2+y^2$. Also the weight enumerator $x^8 + 14x^4y^4 + y^8$ of the $[8,4,4]$ extended binary Hamming code must be invariant since it arises from a plane of order 2; we consider the slightly simpler polynomial

$$\eta_2(x,y) = \tfrac{1}{4}\big[x^8 + 14x^4y^4 + y^8 - \eta_1^4\big] = x^2y^2(x^2-y^2)^2$$

which is evidently also invariant. In order to show that η_1 and η_2 generate the full ring of invariants $\mathbb{C}[x,y]^G$, it suffices to show that η_1 and η_2 are algebraically independent. Suppose that there exists a nonzero polynomial

$$h(T_1,T_2) = cT_1^{i_1}T_2^{i_2} + \cdots \in \mathbb{C}[T_1,T_2]$$

such that $h(\eta_1,\eta_2)=0$. Here $c\neq 0$ and $T_1^{i_1}T_2^{i_2}$ is the lex-highest monomial appearing in $h(T_1,T_2)$: that is, for every monomial $T_1^{j_1}T_2^{j_2}$ appearing in h, either

$$\begin{aligned} &i_1 > j_1, \text{ or} \\ &i_1 = j_1 \text{ and } i_2 \geq j_2\,. \end{aligned}$$

Then the coefficient of $x^{2i_1+6i_2}y^{2i_2}$ in $h(\eta_1,\eta_2)$ is $c\neq 0$; in particular, $h(\eta_1,\eta_2)$ cannot vanish. This shows that η_1 and η_2 are algebraically independent. It follows (cf. Example A6.8) that $\{\eta_1,\eta_2\}$ is a fundamental set of invariants, i.e.

$$\mathbb{C}[x,y]^G = \mathbb{C}[\eta_1,\eta_2].$$

In particular if there exists a projective plane of order $n\equiv 2 \mod 4$, then its extended binary code of length $N+1=n^2+n+2$ must be a polynomial combination of x^2+y^2 and $x^2y^2(x^2-y^2)^2$. A stronger result may be shown: It is not hard to show (Exercise #8) that all codewords in $\widehat{\mathcal{C}}$ have weight divisible by 4, so that $A_{\widehat{\mathcal{C}}}(x,y)$ is invariant under the group $\left\langle G, \left[\begin{smallmatrix}1 & 0\\ 0 & i\end{smallmatrix}\right]\right\rangle$ of order 192. The ring of invariants for this larger group has Hilbert series

$$\frac{1}{(1-t^8)(1-t^{24})}$$

and a set of fundamental invariants is

$$x^8+14x^4y^4+y^8 \qquad \text{and} \qquad x^4y^4(x^4-y^4)^4.$$

Every plane of order $n\equiv 2 \mod 4$ has extended binary code whose weight enumerator is a polynomial combination of the latter two polynomials.

For example this line of reasoning played a role in the proof of the nonexistence of a projective plane of order 10. For a history of this work, see Lam [36]. Early work on a possible plane of order 10 showed that the weight enumerator $A_{\mathcal{C}}(x,y)$ would be uniquely determined if the number of codewords of weight 12, 15 and 16 were known. Some geometric arguments on the part of several researchers (everything from graduate student slave labour to the intensive efforts of Fields Medalist John G. Thompson), and extensive case-checking by computer, showed that no codewords of weight 12, 15 or 16

could exist in such a code, and that the weight enumerators of $\mathcal{C}$ and the extended code $\widehat{\mathcal{C}}$ must have the form

$$\begin{aligned} A_{\mathcal{C}}(x,y) = {} & x^{111} \\ & + 111\,x^{100}y^{11} \\ & + 24{,}675\,x^{92}y^{19} \\ & + 386{,}010\,x^{91}y^{20} \\ & + 18{,}864{,}495\,x^{88}y^{23} \\ & + 78{,}227{,}415\,x^{87}y^{24} \\ & + 2{,}698{,}398{,}790\,x^{84}y^{27} \\ & + 8{,}148{,}873{,}195\,x^{83}y^{28} \\ & + 166{,}383{,}964{,}620\,x^{80}y^{31} \\ & + 415{,}533{,}405{,}150\,x^{79}y^{32} \\ & + 5{,}023{,}148{,}053{,}500\,x^{76}y^{35} \\ & + 10{,}604{,}483{,}511{,}375\,x^{75}y^{36} \\ & + 78{,}347{,}862{,}432{,}300\,x^{72}y^{39} \\ & + 141{,}031{,}595{,}676{,}060\,x^{71}y^{40} \\ & + 653{,}162{,}390{,}747{,}370\,x^{68}y^{43} \\ & + 1{,}009{,}413{,}831{,}402{,}540\,x^{67}y^{44} \\ & + 2{,}982{,}186{,}455{,}878{,}665\,x^{64}y^{47} \\ & + 3{,}976{,}279{,}652{,}851{,}020\,x^{63}y^{48} \\ & + 7{,}582{,}305{,}834{,}092{,}682\,x^{60}y^{51} \\ & + 8{,}748{,}789{,}607{,}170{,}360\,x^{59}y^{52} \\ & + 10{,}841{,}059{,}295{,}003{,}634\,x^{56}y^{55} \\ & + 10{,}841{,}059{,}295{,}003{,}634\,x^{55}y^{56} \\ & + 8{,}748{,}789{,}607{,}170{,}360\,x^{52}y^{59} \\ & \vdots \\ & + y^{111}; \end{aligned}$$

$$\begin{aligned} A_{\widehat{\mathcal{C}}}(x,y) = {} & x^{112} \\ & + 111\,x^{100}y^{12} \\ & + 410{,}685\,x^{92}y^{20} \\ & + 97{,}091{,}910\,x^{88}y^{24} \\ & + 10{,}847{,}271{,}985\,x^{84}y^{28} \\ & + 581{,}917{,}369{,}770\,x^{80}y^{32} \\ & + 15{,}627{,}631{,}564{,}875\,x^{76}y^{36} \\ & + 219{,}379{,}458{,}108{,}360\,x^{72}y^{40} \\ & + 1{,}662{,}576{,}222{,}149{,}910\,x^{68}y^{44} \\ & + 6{,}958{,}466{,}108{,}729{,}685\,x^{64}y^{48} \\ & + 16{,}331{,}095{,}441{,}263{,}042\,x^{60}y^{52} \\ & + 21{,}682{,}118{,}590{,}007{,}268\,x^{56}y^{56} \\ & + 16{,}331{,}095{,}441{,}263{,}042\,x^{52}y^{60} \\ & + 6{,}958{,}466{,}108{,}729{,}685\,x^{48}y^{64} \\ & + 1{,}662{,}576{,}222{,}149{,}910\,x^{44}y^{68} \\ & + 219{,}379{,}458{,}108{,}360\,x^{40}y^{72} \\ & + 15{,}627{,}631{,}564{,}875\,x^{36}y^{76} \\ & + 581{,}917{,}369{,}770\,x^{32}y^{80} \\ & + 10{,}847{,}271{,}985\,x^{28}y^{84} \\ & + 97{,}091{,}910\,x^{24}y^{88} \\ & + 410{,}685\,x^{20}y^{92} \\ & + 111\,x^{12}y^{100} \\ & + y^{112} \end{aligned}$$

The fact that the extended code $\widehat{\mathcal{C}}^{\perp} = \widehat{\mathcal{C}}$ has just 111 codewords of weight 12, arising from the lines of the plane, means (by Exercise #7) that no hyperovals can exist in such a plane; this result itself was a huge computational feat. The final task was to consider all possible configurations for the 24,675 codewords of weight 19 in $\mathcal{C}$ (achieved by Thompson) and finally to show that none of these could lead to a projective plane of order 10. This required thousands of hours of supercomputer time during the late 1980's. The final announcement of the nonexistence of a projective plane of order 10, spread like wildfire, not only through mathematical circles, but also through the popular media. To this day, Lam refers to this as a 'computer result' rather than as a 'proof'. His work gives fuel to the debate over the role of computers in proving mathematical results.

If we wish to determine whether or not there exist planes of order 15, 18, 20, etc. then evidently another method will be required; if we follow the same method used to rule out the plane of order 10, the computer resources required would vastly exceed all the resources available in the world. Clearly, a better idea is needed!

Exercises 13.

1. Find the Smith normal form of the 'projective plane of order 1' with incidence matrix

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

2. Let $B = nI + J$ be an $N \times N$ matrix with 1's off the main diagonal and n+1's on the main diagonal. Let 1 be the $N \times 1$ vector of 1's, and let $1^\perp$ be its orthogonal complement. Show that $\langle 1 \rangle$ and its orthogonal complement $1^\perp$ are eigenspaces for B and determine the corresponding eigenvalues. Hence evaluate the determinant of B.

3. Denote by $\nu(A)$ the **nullity** of a matrix A, i.e. the dimension of the right null space of A. Show that $\nu(AB) \leq \nu(A) + \nu(B)$ assuming the matrix product AB is defined.

4. Let A be the incidence matrix of a projective plane of order n, so that $AA^T = nI + J$. Using Exercises #2,3, show that
   (a) The p-rank of A is n^2+n+1 for any prime p not dividing $n(n+1)$.
   (b) The p-rank of A is $n(n+1)$ for any prime p dividing n+1.
   (c) The p-rank of A is at most $n(n+1)/2$ for any prime p dividing n.

5. Let A be an $m \times n$ matrix, and let $R = A\mathbb{Z}^n$, which is the additive group (i.e. $\mathbb{Z}$-module) generated by the columns of A.
   (a) Show that the quotient group $\mathbb{Z}^m/R$ is unchanged if A is replaced by MAN where M and N are integer matrices with determinant ± 1, having size $m \times m$ and $n \times n$ respectively. That is, show that the quotient group $\mathbb{Z}^m/MAN\mathbb{Z}^n$ is isomorphic to $\mathbb{Z}^m/R$.
   (b) If $d_{11}, \ldots, d_{\mu\mu}$ are the elementary divisors of A where $\mu = \min\{m, n\}$, show that the quotient group
   $$\mathbb{Z}^m/R \cong (\mathbb{Z}/d_{11}\mathbb{Z}) \times (\mathbb{Z}/p_{\mu\mu}\mathbb{Z}).$$
   (c) Conclude that the Smith normal form of A is unique, assuming the Fundamental Theorem of Finitely Generated Abelian Groups: Every finitely generated abelian group G is uniquely expressible as
   $$G \cong (\mathbb{Z}/d_1\mathbb{Z}) \times (\mathbb{Z}/d_r\mathbb{Z})$$
   where the non-negative integers satisfy $d_1 \,|\, d_2 \,|\, \cdots \,|\, d_r$ and r is the minimum number of generators of G.

6. Let A be an $m \times n$ integer matrix. An **integral elementary row operation** on A is any of the following: a permutation of the rows of A, or the addition of an integer multiple of any row to any different row, or the multiplication of any row by -1. We define an integral elementary column operation similarly.
   (a) Using the extended Euclidean algorithm, show that there exists a sequence of elementary row and column operations which when applied to A, yields an $m \times n$ matrix whose $(1,1)$ entry equals the gcd of all entries of A, and the remaining entries in the first row and column are all zero.
   (b) Show that by repeated application of (a) one obtains the Smith normal form of A.
   (c) Use the algorithm suggested above to obtain the Smith normal form for the projective plane of order 2.

7. Let $\mathcal{C}$ be the binary code of a projective plane $(\mathfrak{P}, \mathfrak{L})$ of even order n, and let $\widehat{\mathcal{C}}$ be the extended code of length n^2+n+2.
   (a) Show that $\mathcal{C}$ has minimum weight n+1, and that $\mathcal{C}$ has exactly n^2+n+1 codewords with this minimum weight, whose supports are precisely the lines of $(\mathfrak{P}, \mathfrak{L})$.
   (b) Show that the dual $\widehat{\mathcal{C}}^\perp$ of the extended code, has minimum weight n+2; and that minimum weight codewords in $\widehat{\mathcal{C}}^\perp$ are of just two types: extended lines of $(\mathfrak{P}, \mathfrak{L})$ (see (a)); and hyperovals of $(\mathfrak{P}, \mathfrak{L})$.

(Thus for $n = 2$ the 14 codewords of weight 4 in the extended $[8, 4, 4]$ binary Hamming code arise from the 7 lines and the 7 quadrangles of the plane.)

Hint. Use Exercise #12.4.

8. Let $\widehat{\mathcal{C}}$ be the extended binary code of a projective plane $(\mathfrak{P}, \mathfrak{L})$ of order $n \equiv 2 \mod 4$, so that $\widehat{\mathcal{C}}$ is self-dual of length n^2+n+2. Show that every codeword in $\widehat{\mathcal{C}}$ has weight divisible by 4.

   Hint. Every codeword is a sum (mod 2) of rows of the augmented matrix $G = [A \ \mathbf{1}^T]$. Use induction on the number of rows in such a sum.

14. The Bruck-Ryser Theorem

The celebrated main theorem of this section provides a general criterion for nonexistence of projective planes of certain orders. Apart from the computer proof of the nonexistence of a plane of order 10, this is the only such nonexistence result currently known! It states

14.1 Theorem (Bruck and Ryser [9]). Suppose there exists a projective plane of order $n \equiv 1$ or $2 \mod 4$. Then $n = a^2 + b^2$ for some $a, b \in \mathbb{Z}$.

In practice the fastest way to check whether n is a sum of two squares uses the prime factorization of n; see Lemma 14.8(iii) below. For example none of the integers 6, 14, 21, 22, 30, 33 is a sum of two squares, so none of these can be the order of a projective plane. With the exception of order 6, which was previously settled by the nonexistence of two orthogonal Latin squares of order 6, these results were not previously known. The original proof of Bruck and Ryser (1949) relies on a deep theorem of Hasse and Minkowski. Since its publication, more elementary (but somewhat ad hoc) proofs have become available; see [31] or [13]. We have chosen to present here a proof in the spirit of the original 1949 publication. Although our proof is not self-contained, we feel it is more important for the student to place this result in a larger context (namely, questions of congruence of rational quadratic forms) and to learn a tool that can be applied in a broader range of settings than just the current one (i.e. questions about possible orders of projective planes). As with many algorithmic approaches, this tool can be effectively applied without a full understanding of why it works.

Consider a symmetric $N \times N$ matrix M with rational entries. Then M represents a **rational quadratic form** $Q_M : \mathbb{Q}^N \to \mathbb{Q}$ defined by $Q_M(v) = vMv^T$. Consider a change of basis for $\mathbb{Q}^N$ of the form $v = \widetilde{v}A$ where A is an invertible $N \times N$ matrix over $\mathbb{Q}$. Relative to the new coordinate vector $\widetilde{v}$, the quadratic form becomes

$$Q_M(v) = vMv^T = \widetilde{v}AMA^T\widetilde{v}^T = Q_{AMA^T}(\widetilde{v}).$$

Thus the new symmetric $N \times N$ matrix AMA^T represents the same quadratic form as the original matrix M, relative to a new basis of $\mathbb{Q}^N$. We say that a matrix $\widetilde{M}$ is **rationally congruent** to M, if $\widetilde{M} = AMA^T$ for some invertible $N \times N$ matrix A with rational entries.

Now a projective plane of order $n \geq 2$ (with a prescribed ordering of its N points and N lines, where $N = n^2 + n + 1$) is equivalent to an $N \times N$ matrix A of 0's and 1's such that

(i) $AJ = JA = (n+1)J$; and
(ii) $AA^T = nI + J$

where J is the $N \times N$ matrix with every entry equal to 1. Rewriting (ii) as $AIA^T = nI+J$, we see that the matrices I and $nI+J$ are rationally congruent. This property imposes strong necessary conditions on the matrices I and $nI+J$, and therefore on the integer n itself, as we describe below.

Congruence transformations are reminiscent of similarity transformations: A matrix $\widetilde{M}$ is **similar** to a matrix M, if $\widetilde{M} = AMA^{-1}$ for some invertible matrix A. (Equivalently, M and $\widetilde{M}$ are similar if they represent the same linear transformation $T : \mathbb{Q}^N \to \mathbb{Q}^N$ relative to different bases. Indeed if $w = vM$ then changing bases via the change-of-basis matrix A for both domain and range gives $w = \widetilde{w}A$ and $v = \widetilde{v}A$, so that $\widetilde{w} = \widetilde{v}AMA^{-1}$.) There is in principle a simple criterion to test whether two symmetric rational matrices are similar: the necessary and sufficient condition is that they have the same spectra, i.e. list of eigenvalues with multiplicity. (The condition for similarity is simplified in the case of symmetric matrices since real symmetric matrices are diagonalizable.)

We thus require a similar criterion to determine whether or not two symmetric $N \times N$ rational matrices are congruent. Fortunately such a criterion exists, and it is presented below without proof; the student is referred to [58] for proofs and more explanation. At least one necessary condition is clear: taking determinants we see that congruent matrices $M, \widetilde{M}$ must satisfy $\det \widetilde{M} = c^2 \det M$ for some nonzero $c \in \mathbb{Q}$. Thus we define the **discriminant** of a quadratic form Q, denoted $\operatorname{disc} Q$, to be the determinant of the associated matrix M. We regard the discriminant as defined only to within multiplication by the square of some nonzero rational number, so that $\operatorname{disc} Q$ is a well-defined element of the quotient group $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$, independent of the choice of basis; here $(\mathbb{Q}^\times)^2 < \mathbb{Q}^\times$ is the subgroup consisting of rational squares.

We first recall the **Legendre symbol** defined for an odd prime p, and an integer $a \not\equiv 0 \mod p$:

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & \text{if } a \text{ is a square mod } p, \\ -1, & \text{otherwise;} \end{cases}$$

This symbol has a unique multiplicative extension to all rationals $a = r/s$ in lowest terms with $r, s \in \mathbb{Z}$ and $p \nmid rs$: one simply defines

$$\left(\frac{r/s}{p}\right) = \left(\frac{rs}{p}\right).$$

Note that this symbol is multiplicative in its upper argument: if $a, b \in \mathbb{Q}$ have no factor of p in either the numerator or denominator when written in lowest terms, then

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

We now introduce the **Hilbert symbol** $(a,b)_p$ for an odd prime p and arbitrary nonzero $a,b\in\mathbb{Q}$: write

$$a = a_0 p^s, \quad b = b_0 p^t$$

where $s,t\in\mathbb{Z}$ and the rational numbers a_0, b_0 in lowest terms have no factor of p in either the numerator or the denominator, and define

$$(a,b)_p = \Big(\frac{(-1)^{st} a_0^t b_0^s}{p}\Big) = (-1)^{st(p-1)/2}\Big(\frac{a_0}{p}\Big)^t\Big(\frac{b_0}{p}\Big)^s.$$

Here we have used the elementary fact that $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$; see Proposition A1.3(i). Some explanation of the meaning of the Hilbert symbol is given in remarks found at the end of this section; for now the reader is asked to simply work with this definition. The Hilbert symbol is easily evaluated by the following rules, where $a,b,c\in\mathbb{Q}^\times$ and p is an odd prime:

(14.2) $(b,a)_p = (a,b)_p = (a,c^2b)_p$;

(14.3) $(a,b)_p = 1$ whenever $a,b\in\mathbb{Z}$ with $ab\not\equiv 0 \mod p$;

(14.4) $(ac,b)_p = (a,b)_p(c,b)_p$;

(14.5) $(a,p)_p = \left(\frac{a}{p}\right)$ whenever $a\in\mathbb{Z}$ with $a\not\equiv 0 \mod p$.

A useful identity, easily proved by considering separately the three mutually exclusive cases $p\,|\,n$, $p\,|\,n{+}1$, $p\nmid n(n{+}1)$, is

(14.6) $(n{+}1,n)_p = (n{+}1,-1)_p$; in particular, $(n{+}1,-n)_p = 1$.

Let M be a nonsingular symmetric $N\times N$ matrix with rational entries, and for $k=1,2,\dots,N$ let M_k denote the upper-left $k\times k$ submatrix of M. In particular $M_N = M$ and $\det M_k\neq 0$ for every k. (If $xM_k = 0$ for some $x\in\mathbb{R}^k$ then $\widetilde{x}M\widetilde{x}^T = xM_kx^T = 0$ where $\widetilde{x}\in\mathbb{R}^N$ consists of the vector x followed by $n{-}k$ zeroes, from which $\widetilde{x}=0$, and this forces $x=0$.) For every odd prime p we define

$$c_p(M) = \big(-1,-\det M\big)_p \prod_{1\le k\le N-1}\big(\det M_k\,,\,-\det M_{k+1}\big)_p\,.$$

14.7 Theorem (Hasse and Minkowski). Let M and $\widetilde{M}$ be nonsingular symmetric $N\times N$ matrices over $\mathbb{Q}$. Then M and $\widetilde{M}$ are rationally congruent iff they have the same discriminant (i.e. to within a rational square factor) and $c_p(\widetilde{M}) = c_p(M)$ for every odd prime p.

Before using this result to prove Theorem 14.1, we warm up with an application to 2×2 matrices. In the following, the **squarefree part** of n is the largest divisor of n which is itself not divisible by any square other than 1.

14.8 Lemma. Consider an integer $n \geq 1$. Then the following conditions are equivalent.

(i) $n = a^2 + b^2$ for some $a, b \in \mathbb{Z}$.

(ii) $n = a^2 + b^2$ for some $a, b \in \mathbb{Q}$.

(iii) The squarefree part of n is divisible by no primes $p \equiv 3 \mod 4$.

(iv) $(n, -1)_p = 1$ for every odd prime p.

(v) The matrices $\left[\begin{smallmatrix} n & 0 \\ 0 & n \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ are rationally congruent.

Proof. Let p be an odd prime. If $n = n_0 p^s$ where $p \nmid n_0$ then $(n, -1)_p = \left(\frac{-1}{p}\right)^s$ from which the equivalence of (iii) and (iv) follows immediately. Both matrices in (v) have square determinant, whereas

$$c_p\left(\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]\right) = (-1, -1)_p (1, -1)_p = 1;$$
$$c_p\left(\left[\begin{smallmatrix} n & 0 \\ 0 & n \end{smallmatrix}\right]\right) = (-1, -n^2)_p (n, -n^2)_p = (-1, -1)_p (n, -1)_p = (n, -1)_p$$

so the equivalence of (iv) and (v) follows from Theorem 14.7. Writing $I_2 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, we see that (v) is equivalent to

$$nI_2 = AI_2A^T = AA^T$$

for some $A \in GL_2(\mathbb{Q})$; but any such matrix has the form $A = \left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$ or $\left[\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right]$ for some $a, b \in \mathbb{Q}$ satisfying $a^2 + b^2 = n$, so clearly (v) is equivalent to (ii).

Obviously (i) implies (ii) and the converse is well-known; in fact we show that (iii) implies (i). Write $n = m^2 p_1 p_2 \cdots p_r$ where $m \geq 1$ and the primes $p_1, p_2, \ldots, p_r \equiv 1$ or $2 \mod 4$. It is well-known that $p_j = a_j^2 + b_j^2$ for some $a_j, b_j \in \mathbb{Z}$; now $n = a^2 + b^2$ where the integers a, b are the real and imaginary parts, respectively, of $m(a_1 + b_1 i)(a_2 + b_2 i) \cdots (a_r + b_r i) \in \mathbb{Z}[i]$ where $i^2 = -1$. □

Proof of Theorem 14.1. If there exists a projective plane of order n then by remarks above, the $N \times N$ identity matrix I_N is rationally congruent to the $N \times N$ matrix

$$M = \begin{bmatrix} n+1 & 1 & 1 & \cdots & 1 \\ 1 & n+1 & 1 & \cdots & 1 \\ 1 & 1 & n+1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & n+1 \end{bmatrix}$$

where $N = n^2 + n + 1$. Since M acts as $(N + n)I = (n + 1)^2 I$ on $\langle (1, 1, \ldots, 1) \rangle$ and as nI on $(1, 1, \ldots, 1)^\perp$, we have $\det M = n^{N-1}(n + 1)^2$; see Exercise #13.2. Since $\det M$ is a rational square, the necessary condition on the discriminants is automatically satisfied.

Exactly the same reasoning gives $\det M_k = (n+k)n^{k-1}$ for $k \in \{1, 2, \ldots, N-1\}$. From this it follows that

$$(\det M_k\,, -\det M_{k+1})_p = \begin{cases} (n+k, -n)_p (n+k+1, -1)_p\,, & k \text{ odd}, \\ (n+k+1, -n)_p (n+k, -1)_p (n, -1)_p\,, & k \text{ even}; \end{cases}$$

for example if k is odd then

$$\begin{aligned}(\det M_k\,, -\det M_{k+1})_p &= \big((n+k)n^{k-1}, -(n+k+1)n^k\big)_p = \big(n+k, -(n+k+1)n\big)_p \\ &= (n+k, -n)_p (n+k, n+k+1)_p = (n+k, -n)_p (n+k+1, -1)_p\end{aligned}$$

by (14.6). The case k even is similar. After cancelling duplicate factors we obtain

$$\begin{aligned}\prod_{1 \le k \le N-1} (\det M_k\,, -\det M_{k+1})_p &= (n+1, -n)_p (n+N, -n)_p (n, -1)_p^{(N-1)/2} \\ &= (n, -1)_p^{(N-1)/2}\end{aligned}$$

using (14.6) and the fact that $n+N = (n+1)^2$ is a square. Considering the remaining factor

$$(-1, -\det M)_p = (-1, -(n+1)^2)_p = 1,$$

we obtain

$$c_p(M) = (n, -1)_p^{(N-1)/2} = \begin{cases} (n, -1)_p, & \text{if } n \equiv 1, 2 \mod 4; \\ 1, & \text{otherwise.} \end{cases}$$

The result follows from Theorem 14.7 and Lemma 14.8. □

While it is possible to simplify somewhat the latter proof by first applying a well-chosen congruence transformation to M, we have chosen not to avail ourselves of such tricks in order to demonstrate the success of the straightforward approach; this we believe will best serve the student who may in the future find application for these techniques.

Finally, the critical Theorem 14.7 deserves some explanation, in lieu of an actual proof. We observe that the quadratic forms x^2+y^2 and x^2-y^2 are not rationally congruent because the first is positive definite while the second is not. This argument uses a necessary and sufficient condition for two matrices to be congruent by a *real* change of variable: the number of positive eigenvalues of the associated matrix must agree, as well as the number of negative eigenvalues. The quadratic forms $x^2 + y^2$ and $x^2 + 2y^2$ are congruent by a *real* change of variable, but not by a *rational* change of variable, since the second form represents the number 3, e.g. with $(x, y) = (1, 1)$, whereas the first form does not; this was shown in our proof of Lemma 14.8 using a 'mod 4' argument. Similarly the quadratic forms $x^2 + 2y^2$ and $x^2 + 10y^2$ are not equivalent: the first represents the number 3 for $(x, y) = (1, 1)$, but the second does not represent the number 3 for any choice of $x, y \in \mathbb{Q}$, as can be shown by arguing either 'mod 3' or 'mod 5'. More generally, some *necessary*

conditions for two rational quadratic forms to be equivalent under a *rational* change of variable are that

(i) they are equivalent under a *real* change of variable, and

(ii) for every prime power p^s, a certain necessary condition mod p^s is satisfied.

Condition (ii) is most readily formulated as congruence over the extension $\mathbb{Q}_p \supset \mathbb{Q}$ consisting of *p-adic numbers,* using the discriminant and the invariants $c_p(M)$ defined above. This is analogous to the condition (i) formulated as congruence over the extension field $\mathbb{Q}_\infty := \mathbb{R} \supset \mathbb{Q}$, which is readily checked by counting positive and negative eigenvalues. Thus the necessary conditions for equivalence of two quadratic forms over $\mathbb{Q}$, are neatly expressed as equivalence over $\mathbb{Q}_p$ for all $p \in \{\infty, 2, 3, 5, 7, 11, \ldots\}$. Amazingly, as the Hasse-Minkowski Theorem shows, these necessary conditions are also sufficient! Moreover the same conclusion holds for two rational quadratic forms that are known to be equivalent over $\mathbb{Q}_p$ for *all but possibly one* $p \in \{\infty, 2, 3, 5, 7, 11, \ldots, \}$; this is why the prime 2 can be omitted from our statement of Theorem 14.7. And fortunately so, since congruence of rational quadratic forms over $\mathbb{Q}_2$ requires a slightly more delicate test than for $p \in \{\infty, 3, 5, 7, 11, \ldots\}$. Incidentally, these happy facts do not hold for forms of degree ≥ 3, where the necessary condition of equivalence over $\mathbb{Q}_p$ for every p is not sufficient to guarantee equivalence over $\mathbb{Q}$; see [58, p.44].

The symbol $(a, b)_p$ as defined above for $a, b \in \mathbb{Q}^\times$ and $p \in \{3, 5, 7, 11, \ldots\}$ is a special case of the Hilbert symbol as defined in [58]. The reader interested in completing the connection will note that in our case, $(a, b)_p$ takes the value 1 or -1, according as the equation

$$ax^2 + by^2 = z^2$$

does, or does not, have any nonzero solution $(x, y, z) \in \mathbb{Q}_p^3$.

Exercises 14.

1. A **Hadamard matrix** is an $n \times n$ matrix with entries ± 1 such that $HH^T = nI$, i.e. the rows of H are orthogonal. Interpret this statement as a congruence of rational quadratic forms. What are the resulting restrictions on n? Are these necessary conditions also sufficient?

2. For every $m \geq 1$, let S_m be the set of all integers in $\{1, 2, \ldots, m\}$ which are expressible as a sum of two squares. It is known [39] (see also [2, p.674]) that S_m has asymptotic density zero in the sense that $|S_m|/m \to 0$ as $m \to \infty$. Use this fact to show that asymptotically, the Bruck-Ryser Theorem excludes about half of all positive integers as possible orders of projective planes, in the following sense: If E_m denotes the set of all integers $n \in \{1, 2, \ldots, m\}$ which fail the necessary conditions of Theorem 14.1 (i.e. $n \equiv 1$ or $2 \mod 4$, but n is not a sum of two squares) then the ratio $|E_m|/m \to \frac{1}{2}$ as $m \to \infty$.

15. Difference Sets

Let G be a finite group of order N. A subset $D \subset G$ is a **(planar) difference set** if every *nonidentity* element $g \in G$ is uniquely expressible in the form $g = d_1 d_2^{-1}$ for

some $d_1, d_2 \in D$. An easy counting argument shows that the parameters must satisfy $N = n^2 + n + 1$ where $n = |D| - 1$. The **right cosets** of D in G are the subsets

$$Dg = \{dg : d \in D\} \quad \text{for } g \in G.$$

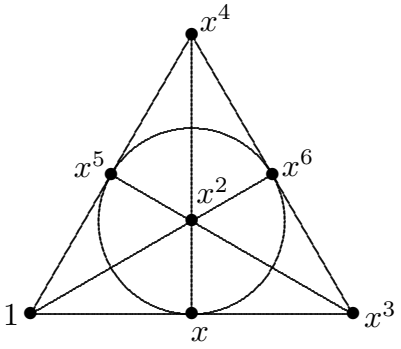
The simplest nontrivial example is given by

$$G = \{1, x, x^2, \ldots, x^6\} \text{ where } x^7 = 1; \quad D = \{1, x, x^3\}.$$

In this case the right cosets of D are the subsets

$$\begin{aligned} D &= \{1, x, x^3\}, \\ Dx &= \{x, x^2, x^4\}, \\ &\vdots \\ Dx^6 &= \{1, x^2, x^6\} \end{aligned}$$

Observe that $\mathfrak{L}_D = \{D, Dx, Dx^2, \ldots, Dx^6\}$ is the line set of a projective plane with point set $\mathfrak{P} = G$. Moreover right-multiplication by x induces a cyclic collineation which regularly permutes both the points and the lines of this plane.



This phenomenon generalizes as follows.

15.1 Theorem. If D is a planar difference set in G, then the right cosets Dg for $g \in G$ form the lines of a projective plane of order $n = |D|-1$ with point set G. Moreover the group G, acting by right-multiplication, induces a collineation group of this plane, which regularly permutes the $N = n^2+n+1$ points and lines of this plane.

Proof. It is clear that the right cosets of D form the blocks of an incidence structure $(G, \mathfrak{L}_D)$ on which G acts by right multiplication, and that G regularly permutes the points of this structure. Moreover there is no loss of generality in assuming that $1 \in D$; otherwise replace D by Dd^{-1} for an arbitrary choice of $d \in D$ (see Exercise #4). To show that this structure is a projective plane, consider any two distinct elements of G. By transitivity of

the action of G, we may assume that these two elements are $\{1, x\}$ where $x \neq 1$. A block Dg contains both 1 and x iff

$$1 = dg \quad \text{and} \quad x = d'g \quad \text{for some } d, d' \in D$$

iff

$$x = d'd^{-1} \quad \text{and} \quad g = d^{-1} \quad \text{for some } d, d' \in D.$$

By definition, there is a unique such ordered pair d, d' in D and hence any two distinct points lie in a unique block. It follows from Exercise #6.3 that $(G, \mathfrak{L}_D)$ is a projective plane of order n. Since G permutes the points transitively, it must permute the lines transitively by Theorem 10.4. But also $|G| = N = n^2+n+1$, so in fact G must permute the lines regularly. □

The following converse of Theorem 15.1 holds:

15.2 Theorem. Let $(\mathfrak{P}, \mathfrak{L})$ be a projective plane of order n admitting a group G of automorphisms permuting the points regularly. Let (P, ℓ) be an arbitrary point-line pair of $(\mathfrak{P}, \mathfrak{L})$, and let D be the set of all $g \in G$ such that P^g lies on ℓ. Then D is a planar difference set in G, and $(\mathfrak{P}, \mathfrak{L})$ is isomorphic to the plane $(G, \mathfrak{L}_D)$ constructed as above from the difference set D.

Proof. By Theorem 10.4, G also permutes the lines of $(\mathfrak{P}, \mathfrak{L})$ regularly. Given any non-identity element $g \in G$, there exists a unique line through both P and P^g, and this line can be written as ℓ^h for some $h \in G$. Since both the points $P^{h^{-1}}$ and $P^{gh^{-1}}$ lie on ℓ, we have $h^{-1} = d$ and $gh^{-1} = d'$ for some $d, d' \in D$; we solve to obtain $g = d'd^{-1}$.

Conversely suppose $g = d'd^{-1}$ for some $d, d' \in D$; then both P and P^g lie on the line $\ell^{d^{-1}}$. The latter line is unique, and since the action of G on lines is regular, the element d^{-1} is uniquely determined; this in turn means that $d' = gd$ is uniquely determined.

Finally let $(G, \mathfrak{L}_D)$ be the plane constructed from the planar difference set D, and consider the mapping $\pi : (\mathfrak{P}, \mathfrak{L}) \to (G, \mathfrak{L}_D)$ defined by acting on points as $P^g \mapsto g$ and on lines as $\ell^h \mapsto Dh$. Since

$$P^g \in \ell^h \quad \text{iff} \quad P^{gh^{-1}} \in \ell \quad \text{iff} \quad gh^{-1} \in D \quad \text{iff} \quad g \in Dh,$$

the map π preserves incidence. Thus π is an isomorphism of projective planes. □

Let $(\mathfrak{P}, \mathfrak{L})$ be a finite projective plane with a collineation group acting regularly on $\mathfrak{P}$, and so also on $\mathfrak{L}$. Fix $P \in \mathfrak{P}$ and $\ell \in \mathfrak{L}$. Write

$$D = \{d \in G : P^d \in \ell\} = \{d \in G : \ell^{d^{-1}} \ni P\}.$$

Consider the map $[-1] : G \to G$, $g \mapsto g^{-1}$; then

$$D^{[-1]} = \{d \in G \,:\, \ell^d \ni P\}.$$

This shows that the while the subset $D \subset G$ is a difference set defining the original plane $(G, \mathfrak{L}_D) \cong (\mathfrak{P}, \mathfrak{L})$, the subset $D^{[-1]} \subset G$ is also a difference set defining the dual plane $(G, \mathfrak{L}_{D^{[-1]}}) \cong (\mathfrak{L}, \mathfrak{P})$. Since *every* planar difference set arises in this way, we have

15.3 Theorem. Let G be a finite group and let $D \subset G$ be any subset. Then D is a planar difference set in G, iff the subset $D^{[-1]} \subset G$ is a planar difference set. In this case the two resulting planes $(G, \mathfrak{L}_D)$ and $(G, \mathfrak{L}_{D^{[-1]}})$ are dual to one another.

Of course if G is abelian then the 'inverse map' $[-1]$ is an automorphism of G mapping D to $D^{[-1]}$, so necessarily the resulting planes $(G, \mathfrak{L}_D)$ and $(G, \mathfrak{L}_{D^{[-1]}})$ are isomorphic in this case.

While a planar difference set D in a group G represents a line in $(G, \mathfrak{L}_D)$, and $D^{[-1]}$ represents a line in the dual plane $(G, \mathfrak{L}_{D^{[-1]}})$, we may ask whether the point set $D^{[-1]} \subset G$ represents anything reasonable in the original plane $(G, \mathfrak{L}_D)$. This question has an interesting answer in the abelian case (see Exercise #5 for a counterexample in the nonabelian case):

15.4 Theorem. Let D be a planar difference set in a finite abelian group G. Then the point set $D^{[-1]} \subset G$ is an oval.

Proof. We have $|D^{[-1]}| = |D| = n+1$ where the plane $(G, \mathfrak{L}_D)$ has order n, so it suffices to show that no three points of $D^{[-1]}$ are collinear. Suppose $x, y, z \in D^{[-1]} \cap Dg$ are distinct and observe that $x^{-1}, y^{-1}, z^{-1}, xg^{-1}, yg^{-1}, zg^{-1} \in D$. Since

$$xyg^{-1} = (xg^{-1})(y^{-1})^{-1} = (yg^{-1})(x^{-1})^{-1},$$

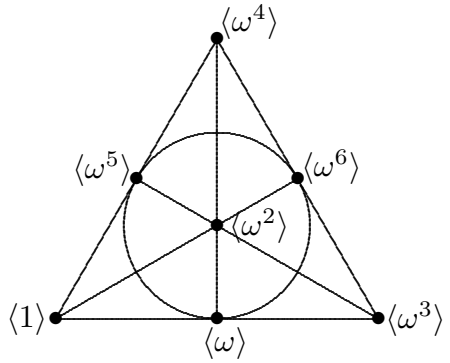
it follows from the definition of a difference set that $xyg^{-1} = 1$. But a similar argument shows that $xzg^{-1} = 1$, which forces $y = z$, a contradiction. □

We now present the standard construction of planar difference sets in cyclic groups of order q^2+q+1 where q is a prime power. These are equivalent to regular collineation groups of finite classical planes. We first illustrate the method by constructing once again a planar difference set in the cyclic group of order 7, which arises from a regular collineation

group on the plane of order 2. Take $E = \mathbb{F}_8 = \mathbb{F}_2[\omega]$ where $\omega^3 = \omega + 1$. The nonzero elements of E are

$$\begin{aligned} &1,\\ &\omega,\\ &\omega^2,\\ &\omega^3 = \omega{+}1,\\ &\omega^4 = \omega^2{+}\omega,\\ &\omega^5 = \omega^3{+}\omega^2 = \omega^2{+}\omega{+}1,\\ &\omega^6 = \omega^3{+}\omega^2{+}\omega = \omega^2{+}1; \end{aligned}$$

note that $\omega^7 = \omega^3{+}\omega = 1$. The seven 2-dimensional subspaces of E over $\mathbb{F}_2$ are the lines of the plane



The map $E \to E$, $x \mapsto x\omega$ is $\mathbb{F}_2$-linear; it regularly permutes the seven points, as well as the seven lines of the plane shown above. All of this generalizes:

15.5 Theorem. Every finite classical plane admits a regular cyclic collineation group.

Proof. Let q be a prime power, and regard $E = \mathbb{F}_{q^3}$ as a 3-dimensional vector space over $F = \mathbb{F}_q$. The points and lines of $\mathbb{P}^2(F)$ may be regarded as the 1- and 2-dimensional subspaces of this vector space. The group $E^\times$ is cyclic; let ω be a generator. The map $E \to E$, $x \mapsto \omega x$ is F-linear and it cyclically permutes the nonzero vectors of E. Let $T \in P\Gamma L_3(F)$ be the corresponding plane collineation, so that T permutes the points of $\mathbb{P}^2(F)$ cyclically and transitively. Since ω^{q^2+q+1} is a primitive $(q-1)$-st root of unity in $E^\times$, it is a generator of $F^\times$ and so T^{q^2+q+1} is the identity collineation, mapping points as $\langle v\rangle \mapsto \langle \omega^{q^2+q+1} v\rangle = \langle x\rangle$. Now the points of $\mathbb{P}^2(F)$, which are the 1-spaces $\langle \omega^i x\rangle$ for $i = 0, 1, 2, \ldots, q^2{+}q$, are cyclically and regularly permuted by the subgroup $\langle T\rangle \le P\Gamma L_2(F)$. The lines must also be permuted regularly, by Theorem 10.4. □

This means that every finite classical projective plane has an incidence matrix which is **circulant**, i.e. the first row of the matrix can be cycled (either left or right) to give the remaining rows of the matrix. In the case of $G = \{1, x, x^2, \ldots, x^6\}$ the difference set

$D = \{1, x, x^3\}$ gives the vector $(1,1,0,1,0,0,0)$ as the first row of an incidence matrix (the characteristic vector of the subset $D \subset G$) and the remaining rows of the incidence matrix are obtained by cycling the first row, thus: $(0,1,1,0,1,0,0)$, $(0,0,1,1,0,1,0)$, etc.

There also exist nonabelian groups admitting planar difference sets; see Exercises #2,3. However, no known difference sets give rise to non-Desarguesian planes. Very little is known about the possibilities for other planar difference sets in non-abelian groups.

The defining properties of a difference set are neatly encoded as identities in the group algebra. Let G be a finite group and consider the algebra of G over $\mathbb{Q}$:

$$\mathbb{Q}G = \Big\{ \sum_{x \in G} a_x x \,:\, a_x \in \mathbb{Q} \Big\}.$$

(See Appendix A3 for a brief summary of the relevant notions used here.) For each integer d, define a map $[d] : G \to G$, $x \mapsto x^d$. This extends to a well-defined map $\mathbb{Q}G \to \mathbb{Q}G$, also denoted by $[d]$:

$$\Big(\sum_{x \in G} a_x x \Big)^{[d]} = \sum_{x \in G} a_x x^d.$$

The latter map is an algebra homomorphism whenever G is abelian; it is an isomorphism whenever G is abelian of order not divisible by d.

15.6 Proposition. Let G be a finite group of order n^2+n+1 where $n \geq 2$, and let $D \subset G$. Define $\delta, \gamma \in \mathbb{Q}G$ by $\gamma = \sum_{x \in G} x$ and $\delta = \sum_{d \in D} d$. Then D is a planar difference set in G iff

$$\delta \delta^{[-1]} = n + \gamma.$$

In this case the relation $\delta\gamma = \gamma\delta = (n+1)\gamma$ also holds.

Proof. This is an immediate consequence of the definitions and the fact that $\alpha\gamma = a\gamma$ for all $\alpha \in \mathbb{Q}G$ where $a \in \mathbb{Q}$ is the sum of the coefficients in α. □

The next result assumes that D is a difference set in an *abelian* group G of order n^2+n+1.

15.7 Hall Multiplier Theorem. Let D be a planar difference set in an abelian group G of order n^2+n+1, and consider any prime p dividing n. Then the map $[p] : G \to G$, $x \mapsto x^p$ permutes the cosets Dg for $g \in G$ and so is a collineation of the projective plane $(G, \mathfrak{L}_D)$ where $\mathfrak{L}_D$ is the set of cosets of D in G. Moreover one of these cosets is fixed by $[p]$.

Proof. Let $\delta = \sum_{x\in D} x$, so that $\delta\delta^{[-1]} = n + \gamma$. We see from the Multinomial Theorem that $\delta^p = \delta^{[p]} + p\alpha$ for some $\alpha \in \mathbb{Z}G$, and so

$$\delta^{[p]}\delta^{[-1]} = \delta^p\delta^{[-1]} - p\alpha\delta^{[-1]} = \delta\delta^{p-1}\delta^{[-1]} + p\alpha_1 = (n+\gamma)\delta^{p-1} + p\alpha_1 = \gamma + p\alpha_2$$

for some $\alpha_1, \alpha_2 \in \mathbb{Z}G$. Since all coefficients on the left side are non-negative, we see that all integer coefficients appearing in α_2 must be non-negative. Multiplying both sides by γ yields $(n+1)^2\gamma = (n^2+n+1)\gamma + p\alpha_2\gamma$, so that $\alpha_2\gamma = (n/p)\gamma$. Applying $[-1]$ yields also $\alpha_2^{[-1]}\gamma = (n/p)\gamma$. Now

$$\begin{aligned}\big(\delta^{[p]}\delta^{[-1]}\big)\big(\delta^{[p]}\delta^{[-1]}\big)^{[-1]} &= \delta^{[p]}\delta^{[-1]}\delta^{[-p]}\delta = \big(\delta\delta^{[-1]}\big)\big(\delta\delta^{[-1]}\big)^{[p]} \\ &= (n+\gamma)(n+\gamma)^{[p]} = (n+\gamma)^2 = n^2 + 2n\gamma + \gamma^2.\end{aligned}$$

On the other hand

$$\begin{aligned}\big(\delta^{[p]}\delta^{[-1]}\big)\big(\delta^{[p]}\delta^{[-1]}\big)^{[-1]} &= (\gamma + p\alpha_2)(\gamma + p\alpha_2^{[-1]}) \\ &= p^2\alpha_2\alpha_2^{[-1]} + p(\alpha_2 + \alpha_2^{[-1]})\gamma + \gamma^2 \\ &= p^2\alpha_2\alpha_2^{[-1]} + 2n\gamma + \gamma^2.\end{aligned}$$

Equating these two expressions yields

$$p^2\alpha_2\alpha_2^{[-1]} = n^2.$$

The expansion $\alpha_2 = \sum_{x\in G} a_x x$ clearly has exactly one nonzero coefficient a_g since the coefficients are non-negative integers and the right side has a single term n^2. It follows that $\alpha_2 = (n/p)g$ for some $g \in G$, and so

$$\delta^{[p]}\delta^{[-1]} = \gamma + ng = \delta\delta^{[-1]}g. \tag{15.8}$$

We claim that

$$\delta^{[p]} = \delta g. \tag{15.9}$$

This will follow from (15.8) if $\delta^{[-1]}$ is invertible in $\mathbb{Q}G$. Let

$$\varepsilon = \tfrac{1}{n} - \tfrac{1}{n(n+1)^2}\gamma \in \mathbb{Q}G;$$

then

$$\delta^{[-1]}\delta\varepsilon = (n+\gamma)\big(\tfrac{1}{n} - \tfrac{1}{n(n+1)^2}\gamma\big) = 1 - \tfrac{1}{(n+1)^2}\gamma + \tfrac{1}{n}\gamma - \tfrac{1}{n(n+1)^2}\gamma^2 = 1.$$

Thus $\delta^{[-1]} \in \mathbb{Q}G$ is invertible and so (15.9) follows.

The cosets of D are permuted by $[p]$ via $Dh \mapsto Dgh^p$ since

$$(\delta h)^{[p]} = \delta^{[p]} h^p = \delta g h^p;$$

thus $[p]$ is a collineation of $(G, \mathfrak{L}_D)$ where $\mathfrak{L}_D$ is the set of cosets of D in G. Note that $[p]$ fixes the point $1 \in G$, so by Corollary 10.3 it fixes some line $Dh \in \mathfrak{L}_D$. □

15.10 Example: Difference Sets in Some Small Groups. Corresponding to the classical plane of order 3 there is a planar difference set in the cyclic group $G = \{1, x, x^2, \ldots, x^{12}\}$ of order 13. In principle one may find a primitive element (an element ω of order $13^3-1 = 2196$) among the nonzero elements of $\mathbb{F}_{2197}$ and proceed as in the example preceding Theorem 15.5. However, the Hall Multiplier Theorem affords us the following shortcut. The orbits of $\sigma = [3] \in \operatorname{Aut} G$ on G are given by

$$\{1\}, \quad \{x, x^3, x^9\}, \quad \{x^2, x^5, x^6\}, \quad \{x^4, x^{10}, x^{12}\}, \quad \{x^7, x^8, x^{11}\}.$$

By Theorem 15.7 there exists a planar difference set D which is a union of orbits of σ. Since $|D| = 4$ we must include $\{1\}$ and one of the orbits of size 3. Any of these will do; take for example $D = \{1, x, x^3, x^9\}$ which we easily verify to be a planar difference set in G.

The planar difference set in the cyclic group of order 7 may be found by similarly enumerating the orbits of $[2]$ on $\{1, x, x^2, \ldots, x^6\}$ where $x^7 = 1$: these orbits are

$$\{1\}, \quad \{x, x^2, x^4\}, \quad \{x^3, x^5, x^6\}$$

and both of the orbits of size 3 give planar difference sets.

Given the difficulty of proving the main questions in the theory of finite projective planes, particularly the question of existence of planes of non-prime power order, one might reasonably hope to make at least some progress in the special case where a regular collineation group exists. Unfortunately to date, such progress has been quite limited. The following result, which is representative of such progress, is included because its proof is short and elegant. It shows that there do not exist projective planes of certain orders $n \in \{6, 10, 12, 14, 15, 18, 21, \ldots\}$ having regular collineation groups.

15.11 Theorem (Wilbrink [66]). Let D be a planar difference set in an abelian group G of order n^2+n+1, and suppose the prime p divides n exactly once. We may assume D is chosen such that $D^{[p]} = D$ by the Hall Multiplier Theorem.

(a) In the group algebra $\mathbb{F}_p G$, the elements $\delta = \sum_{d \in D} d$ and $\gamma = \sum_{x \in G} x$ satisfy

$$\delta^{p-1} + (\delta^{[-1]})^{p-1} = 1 + \gamma.$$

(b) If $p \in \{2, 3\}$ then $n = p$.

Proof. Recall that every $\alpha \in \mathbb{F}_p G$ satisfies $\alpha\gamma = a\gamma$ where $a \in \mathbb{F}_p$ is the sum of the coefficients in α. The following identities clearly hold in characteristic p:

$$\delta\delta^{[-1]} = \gamma; \quad \gamma^2 = \gamma; \quad \delta\gamma = \gamma; \quad \delta^{[-1]}\gamma = \gamma.$$

From the multinomial expansion of δ^p we also have

$$\delta^p = \delta^{[p]} = \delta. \tag{15.12}$$

Since $p \nmid |G|$ the algebra $\mathbb{F}_p G$ is semisimple by Maschke's Theorem A3.10, and so by Theorem A3.16,

(15.13) Every ideal $\mathcal{I} \subseteq \mathbb{F}_p G$ is a principal ideal generated by a unique idempotent $e \in \mathbb{F}_p G$; we write $\mathcal{I} = (e)$.

(15.14) Given ideals $\mathcal{I}_1, \mathcal{I}_2 \subseteq \mathbb{F}_p G$ let $\mathcal{I}_i = (e_i)$ where e_i is the unique idempotent generating $\mathcal{I}_i$; then

$$\mathcal{I}_1 + \mathcal{I}_2 = (e_1{+}e_2 - e_1e_2); \quad \mathcal{I}_1 \cap \mathcal{I}_2 = (e_1e_2)$$

where $e_1{+}e_2{-}e_1e_2$ and e_1e_2 are the unique idempotents generating the ideals $\mathcal{I}_1{+}\mathcal{I}_2$ and $\mathcal{I}_1 \cap \mathcal{I}_2$ respectively.

In particular consider the ideals $\mathcal{I}_1 = (\delta)$ and $\mathcal{I}_2 = (\delta^{[-1]})$. Multiplying (15.12) by δ^{p-2} gives $(\delta^{p-1})^2 = \delta^{p-1}$ so $\delta^{p-1} \in \mathcal{I}_1$ is its unique idempotent generator. Similarly $(\delta^{[-1]})^{p-1} \in \mathcal{I}_2$ is its unique idempotent generator. Therefore the unique idempotent generating $\mathcal{I}_1 \cap \mathcal{I}_2$ is

$$\delta^{p-1}(\delta^{[-1]})^{p-1} = \gamma^{p-1} = \gamma.$$

Thus $\mathcal{I}_1 \cap \mathcal{I}_2$ is 1-dimensional over $\mathbb{F}_p$; also $\mathcal{I}_1 + \mathcal{I}_2$ has as its unique idempotent generator the element

$$\delta^{p-1} + (\delta^{[-1]})^{p-1} - \gamma.$$

By Theorem 13.2 the dimension of $\mathcal{I}_1$ is $\frac{1}{2}\big(n^2{+}n{+}2\big)$. Since $D^{[-1]}$ is also a planar difference set, the dimension of $\mathcal{I}_2$ is also $\frac{1}{2}\big(n^2{+}n{+}2\big)$ and so

$$\dim(\mathcal{I}_1 + \mathcal{I}_2) = \dim \mathcal{I}_1 + \dim \mathcal{I}_2 - \dim(\mathcal{I}_1 \cap \mathcal{I}_2) = n^2{+}n{+}1,$$

i.e. $\mathcal{I}_1 + \mathcal{I}_2 = \mathbb{F}_p G$. Since $1 \in \mathbb{F}_p G$ is the unique idempotent generating $\mathbb{F}_p G$ we have $\delta^{p-1} + (\delta^{[-1]})^{p-1} - \gamma = 1$. This proves (a).

If $p = 2$ then

$$\delta + \delta^{[-1]} = 1 + \gamma = \sum_{x \neq 1} x$$

where the right side has $n^2{+}n$ nonzero coefficients, whereas the left side has at most $2(n{+}1)$ nonzero coefficients. This gives $n(n{+}1) \leq 2(n{+}1)$ and so $n \leq 2$.

If $p = 3$ then

$$\delta^2 + (\delta^{[-1]})^2 = 1 + \gamma = 2 + \sum_{x \neq 1} x. \tag{15.15}$$

Writing $D = \{d_0, d_1, \ldots, d_n\}$ and expanding

$$\delta^2 = \sum_{0 \leq i \leq n} d_i^2 \; + \; 2 \sum_{0 \leq i < j \leq n} d_i d_j, \tag{15.16}$$

the $\binom{n+1}{2}$ terms $d_i d_j$ are distinct since $d_i d_j = d_k d_\ell$ implies $d_i d_k^{-1} = d_\ell d_j^{-1}$. Similarly the $d_i d_j$ terms are distinct from the d_i^2 terms, so (15.16) has exactly $n{+}1$ terms with coefficient 1 and $\binom{n+1}{2}$ terms with coefficient 2. The same holds for the expansion of $(\delta^{[-1]})^2$, so the left side of (15.15) has at most $2(n{+}1) + \binom{n+1}{2}$ terms with coefficient 1. Thus

$$n^2{+}n \; \leq \; 2(n{+}1) + \tbinom{n+1}{2}$$

which yields $n \leq 4$. □

Exercises 15.

1. Construct planar difference sets in the cyclic groups of orders 21 and 31 by the method of Example 15.10.
2. Let G be the nonabelian group of order 21 generated by elements σ, τ subject to the defining relations $\sigma^3 = \tau^7 = 1;\ \tau\sigma = \sigma\tau^2$. Show that $D = \{\tau, \tau^2, \tau^4, \sigma, \sigma^2\}$ is a planar difference set in G. Explain why the resulting projective plane $(G, \mathfrak{L}_D)$ of order 4 must be classical (and therefore isomorphic to the plane constructed from a planar difference set in the *cyclic* group of order 21 as in Exercise #1).
3. The following construction generalizes that of Exercise #2. Consider the field $F = \mathbb{F}_q$ where $q \equiv 1 \mod 3$ and let $E = \mathbb{F}_{q^3}$. Let $\omega \in E$ be a primitive element, i.e. the multiplicative order of ω is $q^3{-}1$. Define F-linear transformations $\sigma, \tau : E \to E$ by $\sigma : x \mapsto x^q;\ \tau : x \mapsto \omega^{3(q-1)}x$. Show that $\langle \sigma, \tau \rangle$ is a nonabelian group of order $q^2{+}q{+}1$ which regularly permutes the 1-dimensional F-subspaces of E.
4. Let D be a planar difference set in a finite group G, and let $a, b \in G$. Show that the subset $aDb = \{adb : d \in D\}$ is also a planar difference set. Show also that the plane $(G, \mathfrak{L}_{aDb})$ is isomorphic to $(G, \mathfrak{L}_D)$.
5. Show by example that the conclusion of Theorem 15.4 does not hold in the nonabelian case.

   Hint. Consider the example $D \subset G$ of Exercise #2, and replace D by an appropriately chosen right coset Dg.

16. Generalized Incidence Matrices

We present a generalization of the concept of a planar difference set, which at the same time generalizes the notion of the incidence matrix for a projective plane. This notion was first introduced by Hughes [29], [30].

Let G be an automorphism group of a finite projective plane $(\mathfrak{P}, \mathfrak{L})$ of order n. Recall that the number of point orbits (call this number w) also equals the number of line orbits. Let $P_1, P_2, \ldots, P_w$ be representatives of the point orbits; and $\ell_1, \ell_2, \ldots, \ell_w$ representatives

of the line orbits. Denote the point stabilizers $\Pi_i = G_{P_i}$ and the line stabilizers $\Lambda_j = G_{\ell_j}$. Set $A_{ij} = \{g \in G : P_i^g \in \ell_j\}$. The desired relations are most easily stated in the group algebra of G over the field of rational numbers. We consider the elements γ, π_i, λ_j, $\alpha_{ij} \in \mathbb{Q}G$ defined by

$$\gamma = \sum G\,; \quad \pi_i = \sum \Pi_i\,; \quad \lambda_j = \sum \Lambda_j\,; \quad \alpha_{ij} = \sum A_{ij}$$

for $i, j \in \{1, 2, \ldots, w\}$. For an arbitrary element $\alpha = \sum_{g\in G} a_g g \in \mathbb{Q}G$ define the **conjugate**

$$\alpha^* = \sum_{g\in G} a_g g^{-1} \in \mathbb{Q}G$$

so that $(\alpha\beta)^* = \beta^*\alpha^*$ for all $\alpha, \beta \in \mathbb{Q}G$. (In Section 15 we denoted $\alpha^{[-1]}$ in place of α^*.) Also denote

$$|\alpha| = \sum_{g\in G} a_g \in \mathbb{Q}$$

so that

$$|\alpha\beta| = |\alpha||\beta|\,; \quad |\alpha \pm \beta| = |\alpha| \pm |\beta|$$

for all $\alpha, \beta \in \mathbb{Q}G$. Thus the map $\mathbb{Q}G \to \mathbb{Q}$, $\alpha \mapsto |\alpha|$ is an algebra homomorphism (usually called the **augmentation map**). Moreover

$$|\gamma| = |G|, \quad |\pi_i| = |\Pi_i|, \quad |\lambda_j| = |\Lambda_j|, \quad |\alpha_{ij}| = |A_{ij}|.$$

Consider the $w \times w$ matrices

$$D_\pi = \operatorname{diag}(\pi_1, \pi_2, \ldots, \pi_w); \quad D_\lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_w);$$
$$A = \big(\alpha_{ij}\big)_{1\le i,j\le w}\,; \quad A^* = \big(\alpha_{ji}^*\big)_{1\le i,j\le w}$$

with entries in $\mathbb{Q}G$ (note that A^* is the **conjugate transpose** of A) and the matrices

$$|D_\pi| = \operatorname{diag}(|\pi_1|, |\pi_2|, \ldots, |\pi_w|); \quad |D_\lambda| = \operatorname{diag}(|\lambda_1|, |\lambda_2|, \ldots, |\lambda_w|);$$
$$|A| = \big(|\alpha_{ij}|\big)_{1\le i,j\le w}\,; \quad J = \big(1\big)_{1\le i,j\le w}$$

with entries in $\mathbb{Z}$. In particular J is the $w \times w$ matrix of 1's.

16.1 Theorem (Hughes [29], [30]**).** Given $G \le \operatorname{Aut}(\mathfrak{P}, \mathfrak{L})$ where $(\mathfrak{P}, \mathfrak{L})$ is a projective plane of order n, define the matrices $D_\pi, D_\lambda, A, |D_\pi|, |D_\lambda|, |A|$ as above. Then

(i) $\operatorname{tr}(|D_\pi|^{-1}) = \operatorname{tr}(|D_\lambda|^{-1}) = \dfrac{n^2+n+1}{|G|}$;

(ii) $J|D_\pi|^{-1}|A| = |A||D_\lambda|^{-1}J = (n+1)J$;

(iii) $A^*|D_\pi|^{-1}A = \gamma J + nD_\lambda$ and $|A^*||D_\pi|^{-1}|A| = |G|J + n|D_\lambda|$; and

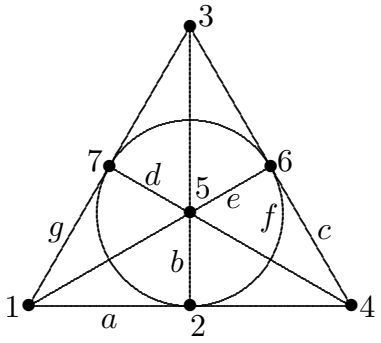
(iv) $A|D_\lambda|^{-1}A^* = \gamma J + nD_\pi$ and $|A||D_\lambda|^{-1}|A^*| = |G|J + n|D_\pi|$.

Before proving these relations let us look at some examples.

16.2 Example: The Trivial Group. If G is the trivial subgroup of $\mathrm{Aut}(\mathfrak{P},\mathfrak{L})$ then $D_\pi = D_\lambda = I_w$ is the identity matrix of order $w = n^2{+}n{+}1$. Moreover A is the usual identity matrix of $(\mathfrak{P},\mathfrak{L})$. In this case (i) asserts that $n^2{+}n{+}1 = n^2{+}n{+}1$; (ii) says that $AJ = JA = (n{+}1)J$; and (iii),(iv) assert that $A^TA = AA^T = nI + J$.

16.3 Example: Regular Collineation Group. If G is a regular collineation group of a plane $(\mathfrak{P},\mathfrak{L})$ then $w = 1$; $D_\pi = D_\lambda = [1]$ and $A = [\delta]$ where δ is arises from a planar difference set, as in Section 15. In this case Theorem 16.1 gives $1 = 1$; $|\delta| = n{+}1$; and $\delta^*\delta = \delta\delta^* = n + \gamma$.

16.4 Example. A more typical situation is illustrated below; here $(\mathfrak{P},\mathfrak{L})$ is the projective plane of order 2, and the collineation group $G = \langle \sigma,\tau : \sigma^2 = \tau^2 = 1,\ \sigma\tau = \tau\sigma^3\rangle$ is dihedral of order 8. The nonidentity elements of G are explicitly listed:



$$\sigma = (1275)(46)(afde)(bg), \qquad \tau = (12)(57)(bg)(ef),$$
$$\sigma^2 = (17)(25)(ad)(ef), \qquad \sigma\tau = \tau\sigma^3 = (25)(46)(ae)(df),$$
$$\sigma^3 = (1572)(46)(aedf)(bg), \qquad \sigma^2\tau = \tau\sigma^2 = (15)(27)(ad)(bg),$$
$$\sigma^3\tau = \tau\sigma = (17)(46)(af)(de).$$

The orbits of G on points and lines are given by

$$1^G = \{1,2,4\}, \qquad a^G = \{a,d,e,f\},$$
$$3^G = \{3\}, \qquad b^G = \{b,g\},$$
$$4^G = \{4,6\}, \qquad c^G = \{c\}.$$

Choosing $1,3,4$ and a,b,c as orbit representatives, we obtain

$$D_\pi = \begin{bmatrix} 1{+}\sigma\tau & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & 1{+}\sigma^2{+}\tau{+}\sigma^2\tau \end{bmatrix}, \qquad D_\lambda = \begin{bmatrix} 1{+}\tau & 0 & 0 \\ 0 & 1{+}\sigma^2{+}\sigma\tau{+}\sigma^3\tau & 0 \\ 0 & 0 & \gamma \end{bmatrix},$$

$$|D_\pi| = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 4 \end{bmatrix}, \qquad |D_\lambda| = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 8 \end{bmatrix};$$

$$A = \begin{bmatrix} 1{+}\sigma{+}\tau{+}\sigma\tau & \sigma{+}\tau{+}\sigma^3{+}\sigma^2\tau & 0 \\ 0 & \gamma & \gamma \\ 1{+}\sigma^2{+}\tau{+}\sigma^2\tau & 0 & \gamma \end{bmatrix}, \qquad |A| = \begin{bmatrix} 4 & 4 & 0 \\ 0 & 8 & 8 \\ 4 & 0 & 8 \end{bmatrix}.$$

We verify the conclusions of Theorem 16.1 in this example as follows:

$$\operatorname{tr}\begin{bmatrix}\frac12&0&0\\0&\frac18&0\\0&0&\frac14\end{bmatrix}=\operatorname{tr}\begin{bmatrix}\frac12&0&0\\0&\frac14&0\\0&0&\frac18\end{bmatrix}=\tfrac78\,;$$

$$\begin{bmatrix}\frac12&0&0\\0&\frac18&0\\0&0&\frac14\end{bmatrix}\begin{bmatrix}4&4&0\\0&8&8\\4&0&8\end{bmatrix}=\begin{bmatrix}2&2&0\\0&1&1\\1&0&2\end{bmatrix}\quad\text{has column sums 3;}$$

$$\begin{bmatrix}4&4&0\\0&8&8\\4&0&8\end{bmatrix}\begin{bmatrix}\frac12&0&0\\0&\frac14&0\\0&0&\frac18\end{bmatrix}=\begin{bmatrix}2&1&0\\0&2&1\\2&0&1\end{bmatrix}\quad\text{has row sums 3;}$$

$$\begin{bmatrix}1+\sigma^3+\tau&0&1+\sigma^2+\tau\\+\sigma^3\tau&&+\sigma^2\tau\\\sigma+\tau+\sigma^3&\gamma&0\\+\sigma^2\tau&&\\0&\gamma&\gamma\end{bmatrix}\begin{bmatrix}\frac12&0&0\\0&\frac18&0\\0&0&\frac14\end{bmatrix}\begin{bmatrix}1+\sigma+\tau&\sigma+\tau+\sigma^3&0\\+\sigma\tau&+\sigma^2\tau&\\0&\gamma&\gamma\\1+\sigma^2+\tau&0&\gamma\\+\sigma^2\tau&&\end{bmatrix}=\begin{bmatrix}\gamma+2(1+\tau)&\gamma&\gamma\\\gamma&\gamma+2(1+\sigma^2&\gamma\\&+\sigma\tau+\sigma^3\tau)&\\\gamma&\gamma&3\gamma\end{bmatrix};$$

$$\begin{bmatrix}4&4&0\\0&8&8\\4&0&8\end{bmatrix}\begin{bmatrix}\frac12&0&0\\0&\frac14&0\\0&0&\frac18\end{bmatrix}\begin{bmatrix}4&0&4\\4&8&0\\0&8&8\end{bmatrix}=\begin{bmatrix}12&8&8\\8&16&8\\8&8&24\end{bmatrix};$$

$$\begin{bmatrix}1+\sigma+\tau&\sigma+\tau+\sigma^3&0\\+\sigma\tau&+\sigma^2\tau&\\0&\gamma&\gamma\\1+\sigma^2+\tau&0&\gamma\\+\sigma^2\tau&&\end{bmatrix}\begin{bmatrix}\frac12&0&0\\0&\frac14&0\\0&0&\frac18\end{bmatrix}\begin{bmatrix}1+\sigma^3+\tau&0&1+\sigma^2+\tau\\+\sigma^3\tau&&+\sigma^2\tau\\\sigma+\tau+\sigma^3&\gamma&0\\+\sigma^2\tau&&\\0&\gamma&\gamma\end{bmatrix}=\begin{bmatrix}\gamma+2(1+\sigma\tau)&\gamma&\gamma\\\gamma&3\gamma&\gamma\\\gamma&\gamma&\gamma+2(1+\sigma^2\\&&+\tau+\sigma^2\tau)\end{bmatrix};$$

$$\begin{bmatrix}4&4&0\\0&8&8\\4&0&8\end{bmatrix}\begin{bmatrix}\frac12&0&0\\0&\frac14&0\\0&0&\frac18\end{bmatrix}\begin{bmatrix}4&0&4\\4&8&0\\0&8&8\end{bmatrix}=\begin{bmatrix}12&8&8\\8&24&8\\8&8&16\end{bmatrix}.$$

Proof of Theorem 16.1. Let T_j be a set of right coset representatives for Λ_j in G; then $|T_j| = [G : \Lambda_j]$ and every line can be uniquely written in the form ℓ_j^t for some $j \in \{1, 2, \dots, w\}$ and $t \in T_j$. Counting lines gives

$$n^2+n+1 = \sum_{1\le j\le w} |T_j| = \sum_{1\le j\le w} |\lambda_j|^{-1}|G|.$$

Dividing both sides by $|G|$ gives one of the relations (i), and the other follows by a dual argument.

Let us count in two different ways the number of elements in the set

$$M_{ij} = \{(g, h) \in G \times G : P_i^g \in \ell_j^h\}.$$

Since $P_i^g \in \ell_j^h$ iff $P_i^{gh^{-1}} \in \ell_j$ iff $gh^{-1} \in A_{ij}$, and every choice of $gh^{-1} \in A_{ij}$ gives rise to $|G|$ pairs $(g,h) \in M_{ij}$, we have $|M_{ij}| = |G||A_{ij}|$. On the other hand the number of $h \in G$ such that $P_i^g \in \ell_j^h$ cannot depend on the choice of $g \in G$, since substituting $h' = hg^{-1}$ for fixed g yields

$$\begin{aligned}|\{h \in G \,:\, P_i^g \in \ell_j^h\}| &= |\{h' \in G \,:\, P_i \in \ell_j^{h'}\}| \\ &= |\Lambda_j||\{t \in T_j \,:\, P_i \in \ell_j^t\}| \\ &= |\Lambda_j||\{\ell \in \ell_j^G \,:\, P_i \in \ell\}|\end{aligned}$$

since each $h' \in G$ lies in a unique coset $\Lambda_j t$ with $t \in T_j$. Thus

$$|M_{ij}| = |G||\Lambda_j||\{\ell \in \ell_j^G \,:\, P_i \in \ell\}|.$$

Equating our two expressions for $|M_{ij}|$ yields

$$|\{\ell \in \ell_j^G \,:\, P_i \in \ell\}| = \frac{|A_{ij}|}{|\Lambda_j|} = \frac{|\alpha_{ij}|}{|\lambda_j|}\,.$$

Since the line set $\mathfrak{L}$ is partitioned into orbits ℓ_j^G for $j = 1, 2, \ldots, w$, we have

$$n+1 = |\{\ell \in \mathfrak{L} \,:\, P_i \in \ell\}| = \sum_{1 \le j \le w} \frac{|\alpha_{ij}|}{|\lambda_j|}\,.$$

Thus the matrix $|A||D_\lambda|^{-1}$ has row sums equal to $n+1$, whence $|A||D_\lambda|^{-1}J = (n+1)J$. The other relation in (ii) follows by a dual argument.

Now fix $i, k \in \{1, 2, \ldots, w\}$ and consider the set

$$N_{ik} = \{(g,h,\ell) \in G \times G \times \mathfrak{L} \,:\, P_i^g\,, P_k^h \in \ell\}.$$

We evaluate in two different ways $\sum gh^{-1} \in \mathbb{Q}G$ where the summation extends over all triples $(g,h,\ell) \in N_{ik}$. Each line $\ell \in \mathfrak{L}$ is uniquely representable in the form $\ell = \ell_j^t$ for some $j \in \{1, 2, \ldots, w\}$ and $t \in T_j$. For this choice of line the corresponding triples (g,h,ℓ_j^t) arise from pairs (g,h) where $g \in A_{ij}t$ and $h \in A_{kj}t$. The sum of gh^{-1} over such pairs (g,h) is $\alpha_{ij}\alpha_{kj}^*$; then varying $t \in T_j$ and $j \in \{1, 2, \ldots, w\}$ gives

$$\sum_{(g,h,\ell) \in N_{ik}} gh^{-1} = \sum_{1 \le j \le w} |T_j|\alpha_{ij}\alpha_{kj}^* = |G| \sum_{1 \le j \le w} |\Lambda_j|^{-1}\alpha_{ij}\alpha_{kj}^*.$$

If $i \neq k$ then every pair $(g,h) \in G \times G$ gives distinct points $P_i^g\,, P_j^h$ which determine a unique line $\ell \in \mathfrak{L}$; in this case

$$\sum_{(g,h,\ell) \in N_{ik}} gh^{-1} = \sum_{g,h \in G} gh^{-1} = |G|\gamma.$$

If $i = k$ then whenever the pair $(g, h) \in G \times G$ satisfies $P_i^g = P_i^h$, there are an additional n lines (i.e. $n+1$ instead of 1) through both P_i^g and P_i^h. Since $P_i^g = P_i^h$ iff $(g, hg^{-1}) \in G \times \Pi_i$, there are $|G||\Pi_i|$ such pairs (g, h), which together contribute a term $n|G|\pi_i$ to our sum. Combining with the previous case gives

$$\sum_{(g,h,\ell)\in N_{ik}} gh^{-1} = |G|\gamma + n\delta_{ik}|G|\pi_i = \begin{cases} |G|\gamma + n|G|\pi_i\,, & \text{if } i = k; \\ |G|\gamma, & \text{otherwise.} \end{cases}$$

Equating our two expressions for $\sum gh^{-1}$ yields

$$\sum_{1\le j\le w} |\lambda_j|^{-1}\alpha_{ij}\alpha_{kj}^* = \gamma + n\delta_{ik}\pi_i\,.$$

This verifies the first relation in (iii), and the second follows by applying the augmentation map $\mathbb{Q}G \to \mathbb{Q},\ \alpha \mapsto |\alpha|$. The proof of (iv) is similar. □

An early application of Hughes' Theorem 16.1 is the following:

16.5 Theorem (Hughes [29], [30]**).** If $(\mathfrak{P}, \mathfrak{L})$ is a projective plane of order $n \equiv 2 \bmod 4$ with $n > 2$, then its automorphism group has odd order.

Thus for example, long before it was determined that a projective plane of order 10 does not exist, it was known that no such plane could have a collineation of order 2. We explain the general idea of the proof of Theorem 16.5: Suppose $(\mathfrak{P}, \mathfrak{L})$ is a plane of order $n \equiv 2 \bmod 4$ with $n > 2$, and suppose there exists $\tau \in \mathrm{Aut}(\mathfrak{P}, \mathfrak{L})$ of order 2. By Theorem 10.7, τ must be an elation. This determines the structure of the matrices A, D_π, etc. for the group $G = \langle\tau\rangle$ of order 2. The second relation of Theorem 16.1 shows that the integer matrices $|D_\pi|^{-1}$ and $2J + n|D_\lambda|$ are rationally congruent. But these matrices are explicitly known in terms of n, and the method of Section 14 yields the required contradiction.

General results comparable to Theorem 16.5 are found only rarely. But Theorem 16.1 is typically applied in many concrete situations as follows. Given a positive integer $n \ge 2$ which satisfies the necessary Bruck-Ryser condition for the possible existence of a projective plane of order n, it is typically nevertheless a very difficult task to decide whether in fact such a plane exists. One may therefore choose one's favourite smallish abstract group, such as S_3 or A_4 or S_4, and then ask whether there exists a projective plane of order n admitting G as a collineation group. (If $n \equiv 2 \bmod 4$ with $n > 2$, then Theorem 16.5 tells us that $|G|$ had better be odd.) Geometric arguments along the lines of Section 10 give some information on the possible types of collineations in G (i.e. their fixed substructures as classified by Theorem 10.5). One thereby enumerates possible rational matrices $|A|$, $|D_\pi|$, $|D_\lambda|$ satisfying the necessary conditions of Theorem 16.1. One then tries to 'lift' these integer matrices A, D_π, D_λ in the group algebra $\mathbb{Q}G$ still satisfying the relations

required by Theorem 16.1. In some cases one obtains a contradiction; in other cases one may find new planes. We have in fact found some new planes of order 25 in this way, the **Wyoming planes** [48].

Exercises 16.

1. Consider the projective plane of order 2, with points and lines labelled as above, and consider the collineation group $G = \langle \sigma, \tau \rangle$ where $\sigma = (346)(257)(aeg)(bdf)$, $\tau = (24)(56)(bc)(df)$. Enumerate the point and line orbits, choosing as orbit representatives the least (i.e. alphabetically first) member in each case, as we did in Example 16.4.

2. Let G be a collineation group of a finite projective plane of order n with w point orbits P_i^G for $i = 1, 2, \ldots, w$; and so also w line orbits ℓ_j^G for $j = 1, 2, \ldots, w$. Consider the product of n^{w-1} with all the orbit sizes:
$$n^{w-1}\Big(\prod_i |P_i^G|\Big)\Big(\prod_j |\ell_j^G|\Big).$$
Show that this integer is a perfect square.

   Examples. Immediately following the statement of Theorem 10.4 we gave an example of a group of order 4 acting on the projective plane of order 2, with 4 point orbits (size 1,1,1,4) and 4 line orbits (size 1,2,2,2). In this case the product becomes $2^3(1{\cdot}1{\cdot}1{\cdot}4)(1{\cdot}2{\cdot}2{\cdot}2) = 2^8$, a perfect square. We leave it to the student to check that the conclusion is satisfied in a couple ready cases: when G is transitive ($w = 1$) and when $G = 1$.

   Hint. Use matrix relations obtained from Theorem 16.1 and take determinants.

17. Blocking Sets

A **blocking set** in a projective plane $(\mathfrak{P}, \mathfrak{L})$ is a set $\mathcal{B} \subset \mathfrak{P}$ of points such that every line meets some point of $\mathcal{B}$. How small can $|\mathcal{B}|$ be and still '*block*' all the lines in this way? Since every superset of a blocking set is also a blocking set, we are primarily interested in **minimal blocking sets**, i.e. blocking sets which do not properly contain any other blocking set.

If $(\mathfrak{P}, \mathfrak{L})$ is a projective plane of order n, it is not hard to see that every blocking set has size $|\mathcal{B}| \geq n{+}1$; for if $P \in \mathfrak{P}$ is a point outside $\mathcal{B}$ then each of the $n{+}1$ lines through P must contain some point of $\mathcal{B}$. We can meet this trivial lower bound by choosing $\mathcal{B}$ to be the point set of some line. We are more interested in a **nontrivial blocking set**, i.e. one which does not contain $[\ell]$ for any line $\ell \in \mathfrak{L}$.

17.1 Proposition (Bruen [10]). If $(\mathfrak{P}, \mathfrak{L})$ is a plane of order n then every nontrivial blocking set has size $|\mathcal{B}| \geq n + \sqrt{n}{+}1$. Equality holds iff $\mathcal{B}$ is the point set of a Baer subplane, in which case n is a square.

Proof. Although the result is due to Bruen, we give a simplified proof due to Blokhuis. If n is a square and $(\mathfrak{P}, \mathfrak{L})$ has a Baer subplane with point set $\mathcal{B}$, then clearly $\mathcal{B}$ is a nontrivial blocking set of size $|\mathcal{B}| = n + \sqrt{n} + 1$.

Suppose $\mathcal{B}$ is a nontrivial blocking set of size $|\mathcal{B}| = n + k$. We must show that $k \geq \sqrt{n} + 1$, with equality iff $\mathcal{B}$ is the point set of a Baer subplane. First observe that every line $\ell \in \mathfrak{L}$ meets $\mathcal{B}$ in at most k points. (For if $\big|[\ell] \cap \mathcal{B}\big| > k$ then choose $P \in \ell$ with $P \notin \mathcal{B}$. One line through P meets $\mathcal{B}$ in at least $k{+}1$ points, and each of the remaining n lines through P meets $\mathcal{B}$ in at least 1 point, so that $|\mathcal{B}| \geq k{+}1{+}n$, a contradiction.)

For each $i \in \{1, 2, \ldots, k\}$, let m_i be the number of lines $\ell \in \mathfrak{L}$ meeting $\mathcal{B}$ in exactly i points. The total number of lines in the plane is

$$\sum_{1 \leq i \leq k} m_i = |\mathfrak{L}| = n^2{+}n{+}1. \tag{17.2}$$

Counting in two different ways, the number of incident point-line pairs $(P, \ell) \in \mathcal{B} \times \mathfrak{L}$ is

$$\sum_{1 \leq i \leq k} i m_i = (n + k)(n + 1). \tag{17.3}$$

Counting in two different ways, the number of pairs of points (P, Q) with $P \in \mathcal{B}$ and $Q \notin \mathcal{B}$ is

$$\sum_{1 \leq i \leq k} i(n + 1 - i)m_i = (n + k)(n^2 + 1 - k); \tag{17.4}$$

this is because there are m_i lines having i choices of $P \in \ell \cap \mathcal{B}$ and $n{+}1{-}i$ choices of $Q \in \ell \smallsetminus \mathcal{B}$. Taking an appropriate linear combination of these relations (namely (17.4), plus $k{-}n$ times (17.3), minus k times (17.2)) gives

$$\sum_{1 \leq i \leq k} (i - 1)(k - i)m_i = n\big[(k{-}1)^2 - n\big]. \tag{17.5}$$

Since the left side of (17.5) is non-negative, we obtain

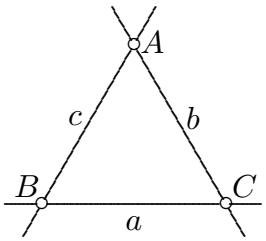
$$k \geq \sqrt{n} + 1.$$

If equality holds then from (17.5) we see that every line $\ell \in \mathfrak{L}$ meets $\mathcal{B}$ in either 1 or $k = \sqrt{n}{+}1$ points. Let $\mathfrak{L}_0$ be the set of lines $\ell \in \mathfrak{L}$ meeting $\mathcal{B}$ in $\sqrt{n}{+}1$ points. Let $P \in \mathcal{B}$, and suppose P lies on r lines of $\mathfrak{L}_0$ and $n + 1 - r$ lines of $\mathfrak{L} \smallsetminus \mathfrak{L}_0$. Counting intersections of lines through P with points of $\mathcal{B}$ gives

$$1 + r\sqrt{n} = n + k = n + \sqrt{n} + 1$$

which yields $r = \sqrt{n} + 1$. Thus $(\mathcal{B}, \mathfrak{L}_0)$ is a 2-$(n{+}\sqrt{n}{+}1,\ \sqrt{n}{+}1,\ 1)$ design embedded in $(\mathfrak{P}, \mathfrak{L})$, i.e. a subplane of order $\sqrt{n}$; see Exercise #6.3. The result follows. □

If n is not a square, how small can a nontrivial blocking set be? It is not hard to construct a nontrivial blocking set with as few as $3n$ points: take three lines a, b, c with no point in common, and remove the three vertices A, B, C of the resulting triangle, as shown:

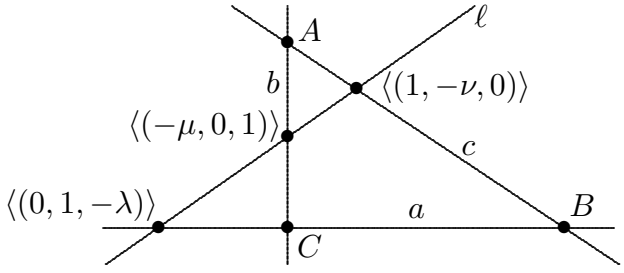


This idea can be improved in the case of classical planes of odd order $n = q$, as follows. Suppose a, b, c are the lines $X = 0$, $Y = 0$ and $Z = 0$ respectively. If a line ℓ does not meet any of the three vertices A, B, C of the triangle (here $A{=}\langle(1,0,0)\rangle$, $B{=}\langle(0,1,0)\rangle$, $C{=}\langle(0,0,1)\rangle$) then ℓ meets a, b, c in points

$$\langle(0,1,-\lambda)\rangle, \quad \langle(-\mu,0,1)\rangle, \quad \langle(1,-\nu,0)\rangle$$

respectively, as shown, where $\lambda\mu\nu \neq 0$. Since these three points are collinear we must have

$$0 = \det \begin{bmatrix} 0 & 1 & -\lambda \\ -\mu & 0 & 1 \\ 1 & -\nu & 0 \end{bmatrix} = 1 - \lambda\mu\nu.$$



In particular either all three or just one of the values λ, μ, ν must be squares in $\mathbb{F}_q$. If we let

$$\mathcal{B} = \{A, B, C\} \cup \big\{\langle(0,1,-\eta)\rangle,\, \langle(-\eta,0,1)\rangle, \langle(1,-\eta,0)\rangle \;:\; \eta \in \mathbb{F}_q \text{ is a nonzero square}\big\}$$

then $\mathcal{B}$ is a blocking set with $|\mathcal{B}| = 3 + 3\big(\frac{q-1}{2}\big) = \frac{3}{2}(q+1)$. This construction is due to di Paola [50], who guessed that this is best possible in the case of $\mathbb{P}^2(\mathbb{F}_p)$, p prime. This guess was verified by Blokhuis [5], using an ingenious polynomial argument:

17.6 Theorem (Blokhuis [5]). A nontrivial blocking set in $\mathbb{P}^2(\mathbb{F}_p)$ has size at least $\frac{3}{2}(p+1)$.

Proof. Let $\mathcal{B}$ be a blocking set in $\mathbb{P}^2(\mathbb{F}_p)$ of size $|\mathcal{B}| = p+k+1$ where $k \le \frac{1}{2}(p+1)$; we must verify the equality $k = \frac{1}{2}(p+1)$. There exists a line meeting $\mathcal{B}$ in just one point (since $|\mathcal{B}| < 2(p+1)$). Choose coordinates (X,Y,Z) for our plane, we may assume that the 'line at infinity' $Z=0$ contains a unique point $\langle(1,0,0)\rangle$ of $\mathcal{B}$, so that

$$\mathcal{B} = \{\langle(1,0,0)\rangle\} \cup \{\langle(a_i,b_i,1)\rangle \,:\, i = 1,2,\dots,p+k\}.$$

Every line not passing through $\langle(1,0,0)\rangle$ has the form $X+uY+tZ=0$ for some $t,u \in \mathbb{F}_p$; and each of these lines must pass through $\langle(a_i,b_i,1)\rangle$ for some i, so the polynomial

$$F(t,u) = \prod_{1\le i\le p+k} (t + a_i + b_i u) \;\in\; \mathbb{F}_p[t,u]$$

vanishes for all values of t,u in $\mathbb{F}_p^2$. Thus

$$F(t,u) = (t^p - t)G(t,u) + (u^p - u)H(t,u)$$

for some $G(t,u), H(t,u) \in \mathbb{F}_p[t,u]$, each of which has degree $\le k$. Taking the $(p+k)$-homogeneous part of each side, we obtain

$$F_0(t,u) = \prod_{1\le i\le p+k} (t + b_i u) \;=\; t^p G_0(t,u) + u^p H_0(t,u)$$

where $G_0(t,u)$ and $H_0(t,u)$ are the k-homogeneous parts of $G(t,u)$ and $H(t,u)$ respectively. Evaluating at $u=1$ gives

$$f = \prod_{1\le i\le p+k} (t + b_i) \;=\; t^p g + h$$

where $f(t) = F_0(t,1),\ g(t) = G_0(t,1),\ h(t) = H_0(t,1)$. For each i we have $(t+b_i) \,\big|\, t^p - t$, so that $t+b_i$ divides $t^p g + h - (t^p - t)g = tg + h$. Write $f(t) = s(t)r(t)$ where $s(t)$ is the product of the *distinct* linear factors of $f(t)$, and $r(t)$ contains all the repeated linear factors. Thus $s \,\big|\, tg+h$ and

$$r \,\big|\, f'(t) = t^p g' + h'.$$

Therefore $f = sr$ divides

$$(tg+h)(t^p g' + h')$$

and so $f = t^p g + h$ also divides

$$(tg+h)(t^p g' + h')g - (tg+h)(t^p g + h)g' \;=\; (tg+h)(gh' - g'h).$$

But $\deg f(t) = p+k$; $\deg(tg+h) = k+1$; and $\deg(gh' - g'h) \le 2k-2$ due to cancellation of the t^{2k-1} terms in gh' and in $g'h$. One possible conclusion from this is that $p+k \le$

$k+1+2k-2$ which yields $k \geq \frac{1}{2}\big(p+1\big)$ as required. The only other possibility is $gh' = g'h$, which yields

$$\frac{d}{dt}\,\frac{h(t)}{g(t)} = \frac{gh' - g'h}{g^2} = 0.$$

Since both $g(t)$ and $h(t)$ have degree less than p, this implies that $h(t)/g(t) = c \in \mathbb{F}_p$, a constant. Thus $h = cg$ and

$$f(t) = (t^p + c)g(t) = (t+c)^p g(t).$$

So p of the b_i's are equal to c. This means that $\mathcal{B}$ contains all points of the line $Y = c$, so that $\mathcal{B}$ is trivial. □

Exercises 17.

1. Show that the projective plane of order two has no nontrivial blocking set.

18. Curves

Let F be a field and let $\overline{F} \supseteq F$ be the algebraic closure of F. Let $f(X,Y,Z) \in F[X,Y,Z]$ be a nonzero homogeneous polynomial of degree $d \geq 1$, so that

$$f(tX,tY,tZ) = t^d f(X,Y,Z)$$

for all $t \in F$. The set of points $\langle(x,y,z)\rangle$ in $\mathbb{P}^2(\overline{F})$ such that $f(x,y,z) = 0$ is a **(projective algebraic) curve** $\mathcal{C}$ of degree d. Those points of the form $\langle(x,y,z)\rangle$ with $x,y,z \in F$ are called the **F-rational points** of the curve. Curves of degree 1 or 2 are of course lines and conics. We say that $\mathcal{C}$ is **irreducible** if $f(X,Y,Z)$ is irreducible over F. *We will assume throughout this section that $\mathcal{C}$ is in fact* **absolutely irreducible**, this being the stronger condition that $f(X,Y,Z)$ is irreducible over $\overline{F}$. In particular, $f(X,Y,Z)$ is uniquely determined by the point set $\mathcal{C}$ up to nonzero scalar multiple.

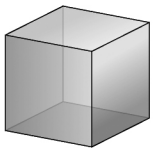
A point $\langle(x,y,z)\rangle \in \mathcal{C}$ is a **singular point** of $\mathcal{C}$ if the three partial derivatives f_x, f_y, f_z all vanish at (x,y,z). The curve $\mathcal{C}$ is **smooth** if none of its points are singular. Many of the basic properties of $\mathcal{C}$ (algebraic, geometric and combinatorial) depend on the *genus* of $\mathcal{C}$. This parameter of the curve $\mathcal{C}$, denoted g, is a non-negative integer whose algebraic definition is given later in this section; but for now we observe that in familiar cases (say $F \in \{\mathbb{Q},\mathbb{R},\mathbb{C}\}$ and $\overline{F} = \mathbb{C}$) the point set $\mathcal{C}$ is a complex curve and therefore a real surface. In the smooth case it is homeomorphic to a '*sphere with g handles*', i.e. one of the surfaces in the list:

$g=0$ (sphere) $g=1$ (torus) $g=2$ $g=3$ etc. ...

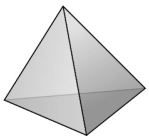
Here we make use of the fact that a smooth complex projective algebraic curve forms a compact connected orientable surface (2-manifold). Each such surface is determined up to homeomorphism by the parameter $g \in \{0, 1, 2, \ldots\}$ and lies in the list above. Note that these are surfaces, not solids; for example when $g = 0$ we have not a solid ball but rather just the sphere which bounds it. Also note that non-orientable surfaces, such as the real projective plane or the Klein bottle, do not arise; the complex analytic structure forces the surface to be orientable. The **Euler characteristic** of a surface Γ is defined by

$$\chi(\Gamma) = V - E + F$$

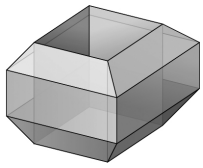
where V, E and F are the number of vertices, edges and faces in a decomposition of Γ into polygonal cells (triangles, squares, etc.). Although the values of V, E and F depend on the decomposition chosen, the value of $\chi(\Gamma)$ is well-defined. Here we illustrate the fact that the 2-sphere S^2 and the torus T^2 have Euler characteristic 2 and 0 respectively:



$$\begin{aligned}\chi(S^2) &= V - E + F \\ &= 8 - 12 + 6 \\ &= 2\end{aligned}$$



$$\begin{aligned}\chi(S^2) &= V - E + F \\ &= 4 - 6 + 4 \\ &= 2\end{aligned}$$



$$\begin{aligned}\chi(T^2) &= V - E + F \\ &= 16 - 32 + 16 \\ &= 0\end{aligned}$$

The genus g may be determined from the Euler characteristic in the case of a sphere with g handles, by the relation

$$\chi(\Gamma) = 2 - 2g.$$

The genus of a curve $\mathcal{C}$ of degree $d \geq 1$ satisfies

$$g \leq \binom{d-1}{2} = \tfrac{1}{2}(d-1)(d-2). \tag{18.1}$$

and equality holds whenever $\mathcal{C}$ is smooth. Thus for example, smooth curves of degree 1 and 2 (that is, lines and conics) have genus 0. Note that a curve of degree zero is just a line

$$\mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$$

where $\mathbb{C}$ is the usual so-called 'complex plane' (which is actually just a *complex affine line*, or a real affine plane, since it has complex dimension 1 and real dimension 2) together with a single additional point ∞ glued to the entire horizon of $\mathbb{C}$. We recognize $\mathbb{P}^1(\mathbb{C}) \simeq S^2$, the **Riemann sphere**, identified with $\mathbb{C} \cup \{\infty\}$ via stereographic projection.

The fact that a nondegenerate conic $\mathcal{C}$ in $\mathbb{P}^2(\mathbb{C})$ is also homeomorphic to S^2 may at first come as a surprise. This fact is a consequence of the *birational equivalence* $f : \mathbb{P}^1(\mathbb{C}) \xrightarrow{\simeq} \mathcal{C}$ which the student may regard simply as a homeomorphism such that both f and f^{-1} are

expressible using rational functions (or polynomials) of the coordinates. This definition requires some delicate interpretation: in our example we may take

$$\begin{aligned}\mathcal{C} &= \{\langle(x,y,z)\rangle \,:\, xz - y^2 = 0,\ (0,0,0) \neq (x,y,z) \in \mathbb{C}^3\} \\ &= \{\langle(s^2, st, t^2)\rangle \,:\, (0,0) \neq (s,t) \in \mathbb{C}^2\}\end{aligned}$$

so that $f : \mathbb{P}^1(\mathbb{C}) \to \mathcal{C}$ is simply

$$\langle(s,t)\rangle \mapsto \langle(s^2, st, t^2)\rangle.$$

The inverse map $f^{-1} : \mathcal{C} \to \mathbb{P}^1(\mathbb{C})$ is defined piecewise by

$$\langle(x,y,z)\rangle \mapsto \begin{cases} \langle(x,y)\rangle, & \text{if } x \neq 0; \\ \langle(y,z)\rangle, & \text{if } z \neq 0. \end{cases}$$

Here it should be noted that x and z cannot vanish simultaneously at any point of the conic $\mathcal{C}$; and in the overlapping region where *both* of the conditions $x \neq 0$ and $z \neq 0$ hold, the two expressions given for $f^{-1}(\langle(x,y,z)\rangle)$ agree. Moreover this overlapping region $xz \neq 0$ includes 'most' of the points of $\mathbb{P}^1(\mathbb{C})$ (technically, it is a Zariski dense subset; but such technicalities we are omitting). A deep result of Hasse and Weil relates the most basic combinatorial properties of $\mathcal{C}$, namely the number of points of the curve with coordinates in a given finite extension $\mathbb{F}_{q^r} \supseteq \mathbb{F}_q$, with the genus of the curve, as follows.

18.2 Hasse-Weil Theorem. Consider a (projective algebraic) curve $\mathcal{C}$ defined over a finite field $\mathbb{F}_q$ by a homogeneous polynomial $f(X,Y,Z)$. Let N_r be the number of $\mathbb{F}_{q^r}$-rational points of $\mathcal{C}$. Then

$$|N_r - (q^r + 1)| \leq 2gq^{r/2}$$

where g is the genus of $\mathcal{C}$.

Note that the projective line $\mathbb{P}^1(\mathbb{F}_{q^r})$ over $\mathbb{F}_{q^r}$ has exactly $q^r + 1$ points. By Theorem 18.2 this is the 'average' or 'typical' number of points on a given curve. For lines (curves of genus 0) this average value is attained exactly. For typical or 'randomly chosen' curves of higher genus (which, by virtue of (29.1) requires higher degree) we typically find that the number of points is distributed rather randomly with a mean value of $q^r + 1$ and standard deviation proportional to $g\sqrt{q^r}$.

Although we do not provide here a proof of the Hasse-Weil Theorem, we mention that it follows from much more general properties of the Zeta-function $Z_{\mathcal{C}}(t)$ of the curve $\mathcal{C}$. That's a Z (upper case Greek letter Zeta), not a Z (upper case Roman z). This function of a complex variable t is given by

$$Z_{\mathcal{C}}(t) = \exp \sum_{r \geq 1} \frac{N_r}{r} t^r \,.$$

The Riemann hypothesis for curves states that $Z_{\mathcal{C}}(t)$ is in fact a rational function of t, i.e. $Z_{\mathcal{C}}(t) \in \mathbb{C}(t)$, of the form

$$Z_{\mathcal{C}}(t) = \frac{(1+\alpha_1 t)(1+\alpha_2 t)\cdots(1+\alpha_{2g} t)}{(1-t)(1-qt)} \tag{18.3}$$

where g is the genus of $\mathcal{C}$ and $\alpha_1, \ldots, \alpha_{2g}$ are algebraic integers of modulus $\sqrt{q}$ which occur in complex conjugate pairs. Statement (18.3) is known as the *Riemann hypothesis for* $\mathcal{C}$. It was first proved in 1973 by Deligne [23], earning him a Fields Medal; for an elementary albeit arduous proof, see [34]. The original Riemann hypothesis concerns the location of the zeroes of the Riemann zeta-function, this being the zeta function of the line over the rational field $\mathcal{Q}$. By the substitution

$$\zeta_{\mathcal{C}}(s) = Z_{\mathcal{C}}(q^{-s})$$

we see that (18.3) implies that the complex zeroes of $\zeta_{\mathcal{C}}$ lie on the line $Re(s) = \frac{1}{2}$; hence the connection with the original Riemann Hypothesis. The assertion (18.3) is one of the original *Weil conjectures,* which more generally give the structure of Zeta functions of varieties defined over finite fields. We show that the Hasse-Weil Theorem is a consequence of (18.3): assuming (18.3),

$$\begin{aligned}
\ln Z_{\mathcal{C}}(t) &= \sum_{1\le i\le 2g} \ln(1+\alpha_i t) - \ln(1-t) - \ln(1-qt) \\
&= \sum_{1\le i\le 2g} \Big(\alpha_i t - \frac{\alpha_i^2 t^2}{2} + \frac{\alpha_i^3 t^3}{3} - \cdots\Big) + \Big(t + \frac{t^2}{2} + \frac{t^3}{3} + \cdots\Big) \\
&\qquad + \Big(qt + \frac{q^2 t^2}{2} + \frac{q^3 t^3}{3} + \cdots\Big)
\end{aligned}$$

and by comparing coefficients of t^r on both sides we obtain

$$N_r = q^r + 1 + \alpha_1^r + \alpha_2^r + \cdots + \alpha_{2g}^r \,. \tag{18.4}$$

Since $|\alpha_i| = \sqrt{q}$ by assumption, we obtain

$$-2gq^{r/2} \;\le\; N_r - (q^r+1) \;\le\; 2gq^{r/2} \tag{18.5}$$

which is the conclusion of Theorem 18.2.

18.6 Example: The Projective Line. Consider a curve of genus 0, i.e. a projective line or conic, say $\mathbb{P}^1 = \mathbb{P}^1(\mathbb{F}_q)$, so that $N_r = q^r + 1$. We obtain

$$\begin{aligned}
\ln Z_{\mathbb{P}^1}(t) &= \sum_{r\ge 1} \big(q^r+1\big)\frac{t^r}{r} \\
&= \sum_{r\ge 1} \frac{t^r}{r} \;+\; \sum_{r\ge 1} \frac{q^r t^r}{r} \\
&= \ln(1-t) + \ln(1-qt)
\end{aligned}$$

using the familiar Taylor expansion for $\ln(1-t)$. Thus

$$Z_{\mathbb{P}^1}(t) = \frac{1}{(1-t)(1-qt)}$$

which is a rational function of the required form.

18.7 Example: The Hermitian Curve. Let $F = \mathbb{F}_q$ where $q = p^{2e}$ and let

$$f(X,Y,Z) \;=\; X^{p^e+1} + Y^{p^e+1} + Z^{p^e+1} \;=\; X^{q_0+1} + Y^{q_0+1} + Z^{q_0+1}$$

where $q_0 = \sqrt{q} = p^e$. (This curve appeared in Section 10 as the set of absolute points with respect to a unitary polarity.) Note that $f_X(X,Y,Z) = (q_0+1)X^{q_0} = X^{q_0} \in \mathbb{F}_q[X,Y,Z]$ and similarly for the other partial derivatives. Since the gradient vector

$$\big(f_X(X,Y,Z),\, f_Y(X,Y,Z),\, f_Z(X,Y,Z)\big) = (X^{q_0}, Y^{q_0}, Z^{q_0})$$

does not vanish at any point of $\mathbb{P}^2(\overline{F}_q)$, the **Hermitian curve** $\mathcal{H}$ consisting of the zeroes of f is smooth; so by (18.1) its genus is

$$g = \binom{q_0}{2} = \tfrac{1}{2}q_0\big(q_0-1\big) = \tfrac{1}{2}\big(q-q_0\big).$$

The Hasse-Weil upper bound for the number of $\mathbb{F}_q$-rational points of $\mathcal{H}$ is

$$N_1 \le q_0^2 + 1 + 2gq_0 = q_0^2 + 1 + q_0^2(q_0-1) = q_0^3 + 1.$$

In fact this upper bound is attained, as we now show. The field F is a quadratic extension of $K = \mathbb{F}_{q_0}$. The norm map

$$N_{F/K} : x \mapsto x^{q_0+1}$$

gives a $(q_0{+}1)$-to-1 map from $F^\times$ to $K^\times$, cyclic groups of order $q_0^2{-}1$ and $q_0{-}1$ respectively. The F-rational points of $\mathcal{H}$ have the form

$$\langle(0,1,a)\rangle \quad \text{and} \quad \langle(1,a,0)\rangle$$

where $N(a) = -1$ ($q_0{+}1$ such values of $a \in F$); also

$$\langle(1,y,z)\rangle$$

where $y \in F$ such that $1 + y^{q_0+1} \neq 0$ (there are $q_0^2 - (q_0+1)$ such values of $y \in F$) and for each such y there are $q_0{+}1$ values of $z \in F$ having the required norm $-(1+y^{q_0+1})$, for a total of

$$N_1 \;=\; 2(q_0+1) + (q_0^2 - q_0 - 1)(q_0+1) \;=\; q_0^3 + 1 \;=\; q^{3/2} + 1.$$

Since the upper bound of (18.5) is attained, for $r = 1$ we obtain

$$\alpha_1 = \alpha_2 = \cdots = \alpha_{2g} = \sqrt{q}$$

and so

$$Z_{\mathcal{H}}(t) = \frac{(1+\sqrt{q}t)^{2g}}{(1-t)(1-qt)} = \frac{(1+\sqrt{q}t)^{q-\sqrt{q}}}{(1-t)(1-qt)}\,.$$

18.8 Example: An Elliptic Curve. An *elliptic curve* is simply a curve of genus 1. Take for example the curve $\mathcal{E}$ defined by

$$f(X,Y,Z) = Y^2Z - X^3 + XZ^2 \in \mathbb{F}_q[X,Y,Z]$$

where $q \equiv 3 \mod 4$. The gradient

$$\big(f_X(X,Y,Z),\, f_Y(X,Y,Z),\, f_Z(X,Y,Z)\big) = \big(Z^2{-}3X^2,\, 2YZ,\, Y^2{+}2XZ\big)$$

cannot vanish at any point of $\mathcal{E}$, so $\mathcal{E}$ is smooth. By (18.1), the genus of $\mathcal{E}$ is 1, i.e. $\mathcal{E}$ is an elliptic curve. It may be shown (see for example [35, pp.59–61]) that

$$N_r = \begin{cases} q^r + 1, & \text{if } r \text{ is odd;} \\ q^r + 1 - 2(-q)^{r/2}, & \text{if } r \text{ is even.} \end{cases}$$

This yields

$$\begin{aligned} \ln Z_{\mathcal{E}}(t) &= -\ln(1-t) - \ln(1-qt) - 2\sum_{\substack{r\geq 1\\ r \text{ even}}} \frac{(-q)^{r/2}}{r}t^r \\ &= -\ln[(1-t)(1-qt)] - \sum_{s\geq 1}\frac{(-qt^2)^s}{s} = -\ln[(1-t)(1-qt)] + \ln(1+qt^2) \end{aligned}$$

so that

$$Z_{\mathcal{E}}(t) = \frac{1+qt^2}{(1-t)(1-qt)}\,.$$

Again, (18.3) holds.

A (nonzero) **differential 1-form** on $\mathcal{C}$ is an expression of the form

$$\omega = \frac{g_1(X,Y,Z)\,dX + g_2(X,Y,Z)\,dY + g_3(X,Y,Z)\,dZ}{h(X,Y,Z)}$$

where $g_1, g_2, g_3 \in \overline{F}[X,Y,Z]$ are homogeneous of the same degree d for some $d \geq 0$, and $h \in \overline{F}[X,Y,Z]$ is homogeneous of degree $d{+}1$. The degree restrictions ensure that

ω does not change under substitutions $(X,Y,Z) \mapsto (\lambda X, \lambda Y, \lambda Z)$ where $\lambda \neq 0$, since $d(\lambda X) = \lambda\, dX$ and similarly for Y and Z. Now consider another form

$$\widetilde{\omega} = \frac{\widetilde{g}_1(X,Y,Z)\, dX + \widetilde{g}_2(X,Y,Z)\, dY + \widetilde{g}_3(X,Y,Z)\, dZ}{\widetilde{h}(X,Y,Z)}\,.$$

We identify $\widetilde{\omega}$ with ω whenever $\widetilde{\omega}$ differs from ω only by a multiple of f or of

$$df = \frac{\partial f}{\partial X}\, dX + \frac{\partial f}{\partial Y}\, dY + \frac{\partial f}{\partial Z}\, dZ\,.$$

To paraphrase this: We can rewrite ω without changing its value, by freely using the relations $f = 0$ and $df = 0$. Since both f and df vanish on $\mathcal{C}$, the expression ω is therefore well-defined at points $\langle (x,y,z) \rangle$ of $\mathcal{C}$. We say ω is **regular at a point** $\langle (x,y,z) \rangle$ of $\mathcal{C}$, if the denominator $h(x,y,z) \neq 0$ (or if ω may be rewritten in some equivalent form, say $\widetilde{\omega}$ as above, with $\widetilde{h}(x,y,z) \neq 0$). We say ω is **regular** (on $\mathcal{C}$) if it is regular at every point of $\mathcal{C}$. In this case it may be necessary to represent ω using different (and of course equivalent) expressions at different points of the curve $\mathcal{C}$, as our examples will illustrate. Denote by $\Omega^1_{\mathcal{C}}$ the vector space (over F) of all regular differential 1-forms on $\mathcal{C}$. Then

(18.9) $\quad \Omega^1_{\mathcal{C}}$ has finite dimension $g \geq 0$.

This gives an algebraic definition of the **genus** g of the curve $\mathcal{C}$.

We verify that this algebraic definition of genus gives the expected answer for those examples we have considered thus far. Consider first the case of a projective line ℓ in $\mathbb{P}^2(F)$ defined by $f(X,Y,Z) = X = 0$; see Example 18.6. In this case both X and dX vanish on the line ℓ, so every nonzero differential 1-form may be uniquely expressed in the form

$$\omega = \frac{g_1(Y,Z)\, dY + g_2(Y,Z)\, dZ}{h(Y,Z)}$$

for some homogeneous $g_i(Y,Z) \in F[Y,Z]$ of degree $d \geq 0$ and $h(Y,Z) \in F[Y,Z]$ of degree $d+1 \geq 1$. We have

$$h(Y,Z) = \prod_{0 \leq i \leq d} (a_i Y + b_i Z)$$

for some $a_i, b_i \in \overline{F}$ with $(a_i, b_i) \neq 0$. We may assume this expression for ω is reduced, i.e. none of the linear factors $a_i Y + b_i Z$ divides the numerator; then ω fails to be regular at the point $\langle (0, b_i, a_i) \rangle$ of $\mathcal{C}$. And no amount of rewriting ω (by including extra X or dX terms) with change this fact. Thus $\Omega^1_{\mathbb{P}^1} = 0$ and $g = 0$, in agreement with our previous finding.

Let us turn to the Hermitian curve $\mathcal{H}$ of Example 18.7 defined by $f(X,Y,Z) = X^{q_0+1} + Y^{q_0+1} + Z^{q_0+1}$. Then both f and

$$df = X^{q_0}\, dX + Y^{q_0}\, dY + Z^{q_0}\, dZ$$

vanish on $\mathcal{H}$. Since

$$X\,df - f\,dX = Y^{q_0}(X\,dY - Y\,dX) + Z^{q_0}(X\,dZ - Z\,dX) = 0$$

on $\mathcal{H}$, we obtain the first equality in the expression

$$\frac{X\,dY - Y\,dX}{Z^{q_0}} \;=\; \frac{Z\,dX - X\,dZ}{Y^{q_0}} \;=\; \frac{Z\,dX - X\,dZ}{Y^{q_0}} \tag{18.10}$$

and the second equality in (18.10) follows by cyclically permuting the variables X, Y, Z. Multiplying the expression (18.10) by $X^iY^jZ^k$ where $i,j,k \geq 0$ and $i{+}j{+}k = q_0{-}2$ gives $\binom{q_0}{2}$ linearly independent differential 1-forms

$$\omega_{ijk} = \frac{X^iY^j}{Z^{q_0-k}}\big(X\,dY - Y\,dX\big) \;=\; \frac{X^iZ^k}{Y^{q_0-j}}\big(Z\,dX - X\,dZ\big) \;=\; \frac{Y^jZ^k}{X^{q_0-i}}\big(Z\,dX - X\,dZ\big).$$

At every point $\langle(x,y,z)\rangle$ of $\mathcal{H}$, at least one of the coordinates x, y, z is nonzero, and so at least one of the expressions for ω_{ijk} listed above has nonvanishing denominator. It may be shown that the ω_{ijk} span $\Omega^1_{\mathcal{H}}$, so that

$$g = \dim \Omega^1_{\mathcal{H}} = \binom{q_0}{2} = \tfrac{1}{2}q_0\big(q_0 - 1\big) = \tfrac{1}{2}\big(q - q_0\big)$$

as previously indicated.

Consider now the elliptic curve $\mathcal{E}$ defined by $f(X,Y,Z) = Y^2Z - X^3 + XZ^2$, introduced in Example 18.8. The defining equation of $\mathcal{E}$ may be rewritten as

$$y^2 = x^3 - x$$

where $x = \frac{X}{Z},\ y = \frac{Y}{Z}$ and so

$$2y\,dy = (3x^2 - 1)\,dx.$$

Consider the differential

$$\omega = \frac{dx}{y} = \frac{2\,dy}{3x^2 - 1}\,,$$

i.e.

$$\omega = \frac{Z\,dX - X\,dZ}{YZ} = 2\,\frac{Z\,dY - Y\,dZ}{3X^2 - Z^2}\,.$$

The only point of $\mathcal{E}$ where both YZ and $3X^2 - Z^2$ vanish is $\langle(0,1,0)\rangle$, so ω is regular at all points of $\mathcal{E}$ except possibly this one. Now rewrite the defining equation of $\mathcal{E}$ in the form

$$w^2x = x^2 - 1$$

where $x = \frac{X}{Z},\ w = \frac{Y}{X}$, so that

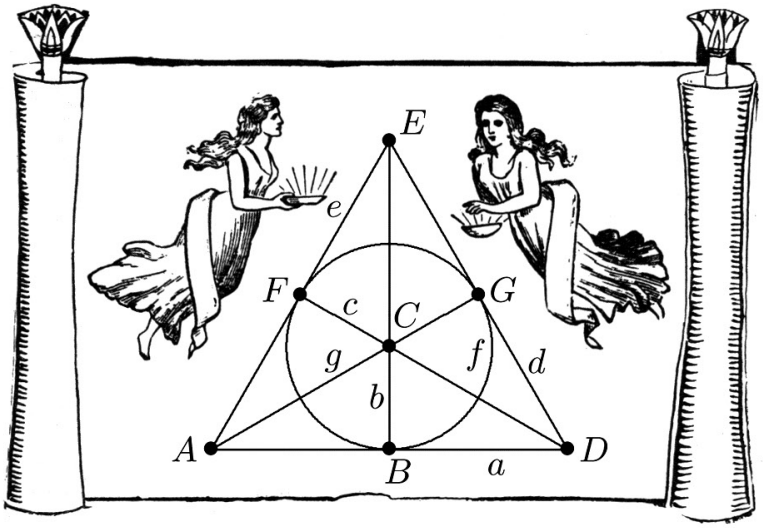
$$\begin{aligned}(2x - w^2)\,dx &= 2xw\,dw;\\ \omega = \frac{dx}{xw} &= \frac{2\,dw}{2x - w^2}\\ &= \frac{2X^2Z\,d(Y/X)}{2X^3 - Y^2Z}\\ &= \frac{2Z(X\,dY - Y\,dX)}{2(Y^2Z + XZ^2) - Y^2Z}\\ &= 2\frac{X\,dY - Y\,dX}{Y^2 + 2XZ}\end{aligned}$$

where the denominator does not vanish at $\langle(0,1,0)\rangle$. Thus in fact ω is regular everywhere. It may be shown that ω spans $\Omega^1_{\mathcal{E}}$, so that the genus of $\mathcal{E}$ is 1 as previously indicated.

Exercises 18.

1. Consider the projective curve $\mathcal{C}$ having the affine description $y^2 + y = x^3$ defined over $\mathbb{F}_2$. The number of $\mathbb{F}_{2^r}$-rational points of $\mathcal{C}$ is
$$N_r = \begin{cases} 2^r + 1, & \text{for } r = 1, 3, 5, \ldots;\\ 2^r + 1 - 2(-2)^{r/2}, & \text{for } r = 2, 4, 6, \ldots.\end{cases}$$
(This includes the one 'point at infinity'.) *You may assume this.*
   (a) Show that $\mathcal{C}$ is smooth.
   (b) What is the genus of $\mathcal{C}$?
   (c) Determine the Zeta function $Z_{\mathcal{C}}(t)$ of $\mathcal{C}$. Express your answer as a rational function of t in the simplest form possible.
2. Consider the curve $\mathcal{C}$ defined in Exercise #1. Find a basis for $\Omega^1_{\mathcal{C}}$.

Part IV



rojective and Polar Spaces

Part IV

Projective and Polar Spaces

19. Classical Affine and Projective Spaces

For every field F and integer $n \geq 0$, we define **classical affine n-space over** F as the incidence system $\mathbb{A}^n(F)$ formed by the affine subspaces of a vector n-space over F. Thus the points, lines, planes, etc. of affine n-space are formed by the cosets of vector k-subspaces for $k = 0, 1, 2, \ldots$ respectively. The field F may be replaced by any skewfield K in this construction, in which case one takes just the subspaces of the *left* vector space K^n over K (or alternatively, the *right* vector space K^n); again, however, this remark does not yield any new finite examples since every finite skewfield is commutative, hence a field. Of more interest to us, and for similar reasons as described in Section 8, are the *projective* spaces which we now introduce.

Let V be an n-dimensional vector space over a field F. The incidence structure formed by the subspaces of V of dimension $k = 1, 2, 3 \ldots, n-1$ form the objects called **points**, **lines**, **planes**, ..., **hyperplanes** of **classical projective** $(n-1)$**-space** over F, denoted $\mathbb{P}V = \mathbb{P}^n(F)$; in general a projective $(k-1)$-subspace[1] of $\mathbb{P}V$ is given by a vector k-subspace of V. The incidence relation between subspaces is the natural inclusion relation. Again, the field F may be replaced by any skewfield K.

In the case of a finite field $F = \mathbb{F}_q$, there are only finitely many vector k-subspaces $U \leq V$ for each $k \in \{1, 2, 3, \ldots, n\}$. One may specify any such subspace as $U = \langle u_1, u_2, \ldots, u_k \rangle$ where $u_1, u_2, \ldots, u_k \in V$ are linearly independent. There are

$$q^n - 1 \text{ choices of } u_1 \in V \setminus \{0\}.$$

Then for each such u_1 there are

$$q^n - q \text{ choices of } u_2 \in V \setminus \langle u_1 \rangle.$$

And for each choice of u_1, u_2 there are

$$q^n - q^2 \text{ choices of } u_3 \in V \setminus \langle u_1, u_2 \rangle;$$

etc. Thus the number of possible ordered lists of k linearly independent vectors $u_1, u_2, \ldots, u_k \in V$, giving an ordered basis for some vector k-subspace of V is

$$(q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{k-1}).$$

A similar count shows that the number of choices of ordered basis for a given k-subspace $U \leq V$ is

$$(q^k - 1)(q^k - q)(q^k - q^2) \cdots (q^k - q^{k-1}).$$

[1] The *projective dimension* of a subspace $\mathbb{P}U \subseteq \mathbb{P}V$, is one less than the vector space dimension of $U \leq V$. Unless we specify projective dimension, then we refer to the usual vector space dimension. For example if V is a 3-dimensional vector space, then its projective dimension is 2, i.e. $\mathbb{P}V$ is a projective 2-space, otherwise known as a projective plane.

Therefore the number of vector k-subspaces of the n-space V is given by the **Gaussian coefficient**

$$\begin{bmatrix} n \\ k \end{bmatrix} = \begin{bmatrix} n \\ k \end{bmatrix}_q := \frac{(q^n-1)(q^n-q)(q^n-q^2)\cdots(q^n-q^{k-1})}{(q^k-1)(q^k-q)(q^k-q^2)\cdots(q^k-q^{k-1})}\,.$$

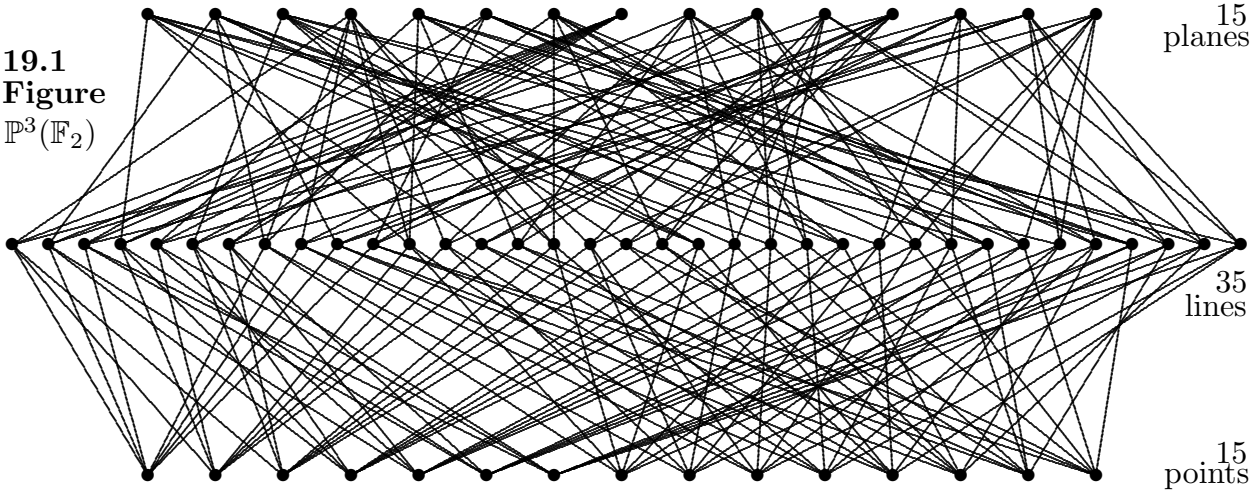
For example projective 3-space over $\mathbb{F}_q$, denoted $\mathbb{P}^3(\mathbb{F}_q)$, has

$$\begin{bmatrix} 4 \\ 1 \end{bmatrix}_q = \frac{q^4-1}{q-1} = (q^2+1)(q+1) \text{ points;}$$

$$\begin{bmatrix} 4 \\ 2 \end{bmatrix}_q = \frac{(q^4-1)(q^4-q^2)}{(q^2-1)(q^2-q)} = (q^2+1)(q^2+q+1) \text{ lines; and}$$

$$\begin{bmatrix} 4 \\ 3 \end{bmatrix}_q = \frac{(q^4-1)(q^4-q^2)(q^4-q^3)}{(q^3-1)(q^3-q)(q^3-q^2)} = (q^2+1)(q+1) \text{ planes.}$$

In the case of the field $\mathbb{F}_2$, projective 3-space has 15 points, 35 lines, and 15 planes. We may represent this geometry by the *Hasse diagram*



19.1 Figure $\mathbb{P}^3(\mathbb{F}_2)$

in which the bottom row lists the 15 points in lexicographic order

$$\langle(0,0,0,1)\rangle,\ \langle(0,0,1,0)\rangle,\ \langle(0,0,1,1)\rangle,\ \langle(0,1,0,0)\rangle,\ \ldots,\ \langle(1,1,1,1)\rangle;$$

the top row lists the 15 planes in lexicographic order

$$\langle(0,0,0,1)^{\mathrm{T}}\rangle,\ \langle(0,0,1,0)^{\mathrm{T}}\rangle,\ \langle(0,0,1,1)^{\mathrm{T}}\rangle,\ \langle(0,1,0,0)^{\mathrm{T}}\rangle,\ \ldots,\ \langle(1,1,1,1)^{\mathrm{T}}\rangle.$$

Here we denote the plane

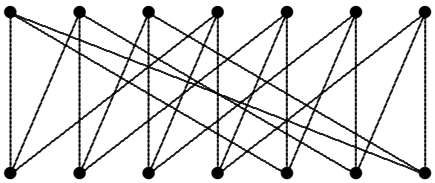
$$\left\{ \langle(x_0\ x_1\ x_2\ x_3)\rangle \ :\ \begin{pmatrix} x_0 & x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 = 0 \right\}$$

by the 1-space $\langle (a_0, a_1, a_2, a_3)^{\mathrm{T}} \rangle$ of column vectors; this notation closely parallels the notation of Section 6 for coordinatizing points and lines of $\mathbb{P}^2(F)$. The middle row of Figure 19.1 lists the 35 lines in lexicographic order of their *Plücker coordinates* (these coordinates for lines will be defined in Section 22). Edges in this graph indicate incidences between points and lines, and between lines and planes. For clarity, we have not shown the incidences between points and planes; but these are implied by the edges shown (for example the first point $P = \langle (0,0,0,1) \rangle$ is incident with the second plane $\pi = \langle (0,0,1,0)^{\mathrm{T}} \rangle$ as shown by the three paths of length 2 in the graph above; these three paths correspond to the three lines ℓ incident with both P and π). Note that two distinct lines ℓ, m may be related in either of two ways: either they meet in a point (in which case they also lie in some plane); or they are disjoint, in which case we say ℓ and m are **skew**. However one does not speak of two distinct lines as being incident; recall that incidence is inclusion, and no line can be contained in another unless they are equal. In $\mathbb{P}^3(\mathbb{F}_2)$,

every point is incident with 7 lines and 7 planes;
every line is incident with 3 points and 3 planes; and
every plane is incident with 7 points and 7 lines.

More than this, the subgraph of Figure 19.1 consisting of all points and lines incident with a given plane, is of course just the incidence graph of the projective plane $\mathbb{P}^2(\mathbb{F}_2)$:

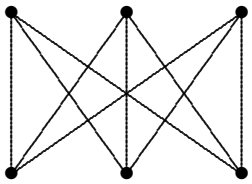
19.2 Figure
Incidence Graph of $\mathbb{P}^2(\mathbb{F}_2)$



Similarly, the subgraph of Figure 19.1 consisting of all lines and planes incident with a given point, also gives the incidence graph of the projective plane $\mathbb{P}^2(\mathbb{F}_2)$ shown in Figure 19.2. To see why this should be true, observe that for every point, i.e. 1-space $P < V$, there is a one-to-one correspondence between subspaces of V containing P, and subspaces of the vector 3-space V/P. Under this correspondence, the lines and planes containing P correspond to points and lines of V/P.

The subgraph of Figure 19.1 consisting of all points and planes incident with a given line ℓ, gives the incidence graph shown in Figure 19.3. This incidence structure is known as a *generalized digon of order* 2. (The explanation for this name will be given in Section 29.) Here each of the three points on ℓ is incident with each of the three planes containing ℓ.

19.3 Figure
Incidence Graph of a Generalized Digon of order 2



A **generalized digon of order** n is an incidence structure with $n+1$ points and the same number of blocks, such that every point is incident with every block; its incidence graph is therefore a complete bipartite graph $K_{n+1,n+1}$.

The space $\mathbb{P}^n(F)$ is self-dual. In this case duality interchanges points with hyperplanes, lines with subspaces of codimension 2; and in general, vector k-subspaces of V with vector $(n{+}1{-}k)$-subspaces. (See Exercise #4.)

A hyperplane of $\mathbb{P}^n(F)$ is isomorphic to $\mathbb{P}^{n-1}(F)$; by deleting this hyperplane, one is left with an affine space $\mathbb{A}^n(F)$. Conversely, if one starts with an affine space $\mathbb{A}^n(F)$ and adds a 'point at infinity' for every parallel class of lines of the affine space (also naturally defining other subspaces at infinity as natural subsets of these new points), one obtains the projective space $\mathbb{P}^n(F)$ as the **projective completion** of the original affine space. This generalizes the process of completing an affine plane to a projective plane as described in Section 7. To see this, note that the points of $\mathbb{P}^n(F)$ are 1-spaces $\langle(x_0, x_1, x_2, \ldots, x_n)\rangle$, partitioned into those points with $x_0 = 0$ (this being a typical hyperplane, a projective $(n{-}1)$-subspace) and those with $x_0 \neq 0$ (which may be normalized by the scalar factor x_0^{-1} so as to be uniquely expressible in the form $\langle(1, a_1, a_2, \ldots, a_n)\rangle$ with $a_1, a_2, \ldots, a_n \in F$ (and the latter points form an affine n-space). In the case $F = \mathbb{F}_q$ this partition of the points of $\mathbb{P}^n(\mathbb{F}_q)$ gives rise to the identity

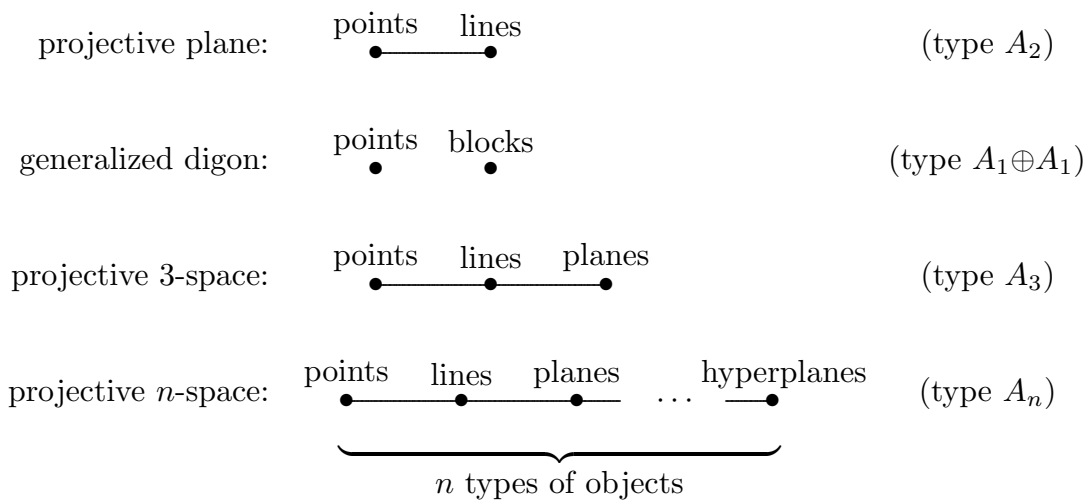
$$\underbrace{q^n + q^{n-1} + \cdots + q + 1}_{\begin{bmatrix} n+1 \\ 1 \end{bmatrix}_q \text{ points of } \mathbb{P}^n(\mathbb{F}_q)} = \underbrace{q^n}_{\substack{\text{points of} \\ \mathbb{A}^n(\mathbb{F}_q)}} + \underbrace{(q^{n-1} + \cdots + q + 1)}_{\begin{bmatrix} n \\ 1 \end{bmatrix}_q \text{ points of } \mathbb{P}^{n-1}(\mathbb{F}_q)} .$$

A homogeneous polynomial $f(X_0, X_1, \ldots, X_n) \in F[X_0, X_1, \ldots, X_n]$ of degree d yields a well-defined point set consisting of all points $\langle(x_0, x_1, \ldots, x_n)\rangle$ such that $f(x_0, x_1, \ldots, x_n) = 0$ (a **hypersurface of degree** d); here we may take the coordinates $x_i \in F$ (which gives the so-called F-rational points of the surface) or more generally, $x_i \in \overline{F}$, the algebraic closure of F. The same processes of homogenization and dehomogenization as described in Section 7 for the case $n = 2$, allow us to form the projective completion of hypersurfaces in affine n-space, and the affine parts of hypersurfaces in projective n-space. For example the surfaces $x^2 + y^2 + z^2 = 1$ and $x^2 = 1 + y^2 + z^2$ in affine 3-space have the same projective completion: the first equation homogenizes to $X^2 + Y^2 + Z^2 = W^2$, which then dehomogenizes (by setting $Z = 1$) to $x^2 + y^2 + 1 = w^2$.

Every semilinear transformation of $V = F^n$ maps k-dimensional subspaces to k-dimensional subspaces, giving a collineation (i.e. automorphism) of the projective geometry. In fact, every collineation of $\mathbb{P}^{n-1}(F)$ is induced by a semilinear transformation. Thus the full collineation group of $\mathbb{P}^{n-1}(F)$ is isomorphic to $P\Gamma L_n(F)$. An **ordered frame** in $\mathbb{P}^{n-1}(F)$ is an ordered $(n{+}1)$-tuple $(P_0, P_1, \ldots, P_n)$ of points, no n of which lie in a hyperplane. We state (but do not prove):

19.4 Fundamental Theorem of Projective Geometry. The full collineation group of $\mathbb{P}^{n-1}(F)$ is isomorphic to $P\Gamma L_{n-1}(F)$. Its normal subgroup $PGL_n(F)$ is regular (i.e. sharply transitive) on ordered frames.

We associate to projective planes, generalized digons, projective 3-spaces, and general projective spaces the following diagrams respectively:



The extent of the association between these **Coxeter-Dynkin diagrams** and the corresponding classes of geometries, is too far-reaching to completely describe here; but a few observations will serve to illustrate the nature of this association. The symmetry of the A_n diagram for $n \geq 2$, which interchanges the nodes representing k-spaces with $(n{+}1{-}k)$-spaces, illustrates the fact that the dual of a projective n-space is again a projective n-space. Given an object U in a geometry Γ (say), the geometry induced on all objects of Γ incident with U is called the **residual geometry** of U, or simply the **residue** of U. Its diagram is obtained from the diagram for Γ by removing the node corresponding to the type of U, together with all edges from that node. For example deleting either of the end nodes of the A_3 diagram gives the A_2 diagram; this illustrates the fact that objects in $\mathbb{P}^3(F)$ incident with a given point or plane, have the structure of a projective plane. Deleting the middle node of the A_3 diagram leaves the $A_1 \oplus A_1$ diagram; this illustrates the fact that the points and planes of $\mathbb{P}^3(F)$ incident with a given line, have the structure of a generalized digon as described above. More generally given an n-space V with a k-subspace $U \leq V$, the residue of U has diagram $A_{k-1} \oplus A_{n-k-1}$ (assuming appropriate allowances are made for the special cases $k \in \{1, n\}$). The objects of this geometry are the subspaces of U, forming the subgeometry $\mathbb{P}U \simeq \mathbb{P}^{k-1}(F)$ of type A_{n-1}, and the subspaces of V containing U. The latter subspaces correspond to subspaces of V/U, which therefore form a subgeometry isomorphic to $\mathbb{P}(V/U) \simeq \mathbb{P}^{n-k-1}(F)$. Moreover every object in the A_{k-1}-subgeometry is incident with every object in the A_{n-k-1}-subgeometry, as is represented by the fact that there are no edges between the corresponding subgraphs in the $A_{k-1} \oplus A_{n-k-1}$ diagram.

Exercises 19.

1. Show that the number of affine k-subspaces of $\mathbb{A}^n(\mathbb{F}_q)$ is $q^{n-k}\begin{bmatrix} n \\ k \end{bmatrix}_q$.
2. Show that the limit of the Gaussian coefficient is $\lim_{q\to 1} \begin{bmatrix} n \\ k \end{bmatrix}_q = \binom{n}{k}$, the binomial coefficient. We regard projective $(n{-}1)$-space as the analogue of an $(n{-}1)$-simplex, which is the incidence system formed

by all subsets of an n-set. Note that a projective plane is thus an analogue of a 2-simplex or triangle, which is the unique 2-$(3, 2, 1)$ design. Also under this analogy, the automorphism group $PGL(n, F)$ of projective $(n-1)$-space is the analogue of the symmetric group S_n , which is the automorphism group of the $(n-1)$-simplex. The $(n-1)$-simplex is often called **projective $(n-1)$-space over the field of order** 1, (even though there is no field of order 1) or **thin projective $(n-1)$-space**, while the other projective $(n-1)$-spaces are **thick**. Compare with comments in Exercise #6.3.

3. Show that the points and hyperplanes of projective $(n-1)$-space over $\mathbb{F}_q$ form a 2-(v, k, λ) design where

$$v = \begin{bmatrix} n \\ 1 \end{bmatrix}_q ; \quad k = \begin{bmatrix} n-1 \\ 1 \end{bmatrix}_q ; \quad \lambda = \begin{bmatrix} n-2 \\ 1 \end{bmatrix}_q .$$

Remarks. Again letting $q \to 1$, this gives a 2-$(n, n-1, n-2)$ design, i.e. an $(n-1)$-simplex. It has n points, and its blocks (i.e. 'hyperplanes') are the $(n-1)$-subsets.

4. Let V be an n-dimensional vector space over a field F, and let V^* be the vector space of all linear functionals $V \to F$; thus V^* is also n-dimensional. Show that there is a one-to-one correspondence between subspaces of V and subspaces of V^* which maps each $U \leq V$ to its *annihilator*

$$U^\circ = \{\phi \in V^* \,:\, \phi(U) = 0\}.$$

Show moreover that $\dim(U^\circ) = n - \dim U$ and that the map $U \mapsto U^\circ$ reverses inclusion: we have $U \leq W \leq V$ iff $U^\circ \geq W^\circ$. Conclude that $\mathbb{P}^{n-1}(F)$ is self-dual.

Hint. Represent elements of V and V^* as row and column vectors of length n, respectively. To evaluate a linear functional at a vector, is simply to right-multiply a row vector by a column vector. Imitate the proof of Theorem 6.5.

5. Let H be the $n \times (2^n-1)$ matrix whose columns are the nonzero vectors in $\mathbb{F}_2^n$, i.e. the points of $\mathbb{P}^{n-1}(\mathbb{F}_2)$. Show that H is the parity check matrix for a $[2^n-1, 2^n-n-1, 3]$ binary code. (See Section 13 for the case $n = 3$.) Verify that this is a perfect 1-error correcting code.

Remark. This is the family of **binary Hamming codes**.

20. Axioms

At this point the reader may wonder whether the concept of a projective space should really be defined axiomatically in terms of natural incidence properties, as we did for projective planes; and *then* the classical examples constructed as we have done above, starting with a vector space over an arbitrary field (or skewfield). However every projective space of projective dimension $n \geq 3$ is necessarily classical. For example here are some natural candidates as axioms for projective 3-space, in which the objects are points, lines and planes; and where three incidence relations are given (between points and lines; between points and planes; and between lines and planes).

(S1) Any two distinct points are incident with exactly one line.
(S2) Any two distinct planes meet in exactly one line.
(S3) Given any line ℓ and any point P not on ℓ, there exists a unique plane incident with both P and ℓ.
(S4) Every plane incident with a given line ℓ is also incident with every point on ℓ.
(S5) Any two distinct lines meet in a point, iff they lie in a common plane.
(S6) There exists a set of five points, of which no four lie in a common plane.

Axiom (S3) asserts that every line or plane may be identified with the set of its points, in such a way that incidence of a line and plane is the same thing as inclusion of the corresponding point sets. These potential axioms are somewhat redundant (see Exercise #20.1). However they do include all that we naturally require of a projective 3-space. In particular these axioms imply that for every plane π, the incidence structure formed by the points and lines on π is exactly that of a projective plane; and dually, for every point P, the incidence structure formed by the lines and planes containing P is also that of a projective plane.

It is possible to axiomatize projective n-space in a style similar to that above, having projective k-subspaces for $k = 0, 1, 2, \ldots, n$ as undefined concepts, and with an incidence relation between subspaces of dimensions j and k for all $j \neq k$, and introducing axioms similar to (S1)–(S6) above. An alternative approach, which we prefer, is the following approach of Veblen and Young (1910).

A **projective n-space** is a linear space $(\mathfrak{P}, \mathfrak{L})$ satisfying the following axioms, where we define a **subspace** to be any subset of $\mathfrak{P}$ which contains every line through two of its points.

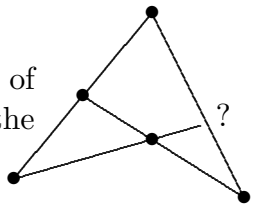
(V1) If P and Q are distinct points then there is exactly one line containing both P and Q.

(V2) There exist three noncollinear points.

(V3) Every line has at least three points.

(V4) The maximum length of a chain of subspaces is $n + 2$.

(V5) **(Veblen-Pasch Axiom)** Every line meeting two sides of a triangle, but none of its vertices, must also meet the third side.



Axiom (V1) is included in the definition of a linear space but we have listed it again for emphasis. In (V4), a **chain** of subspaces of length n has the form $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_n$ where each $\mathcal{S}_i$ is a subspace. For example in projective 3-space every maximal chain has the form

$$\emptyset \subset \{P\} \subset \ell \subset \pi \subset \mathfrak{P}$$

where P is a point lying on a line ℓ in a plane π. Every subspace $\mathcal{S}$ has a well-defined projective dimension $k \geq -1$, where $k+2$ is the maximum length of any chain of subspaces contained in $\mathcal{S}$. Here the empty subspace $\emptyset$ has projective dimension -1, points have projective dimension 0, lines have projective dimension 1, planes have projective dimension 2, etc. with the full point set $\mathfrak{P}$ being the unique subspace of projective dimension n.

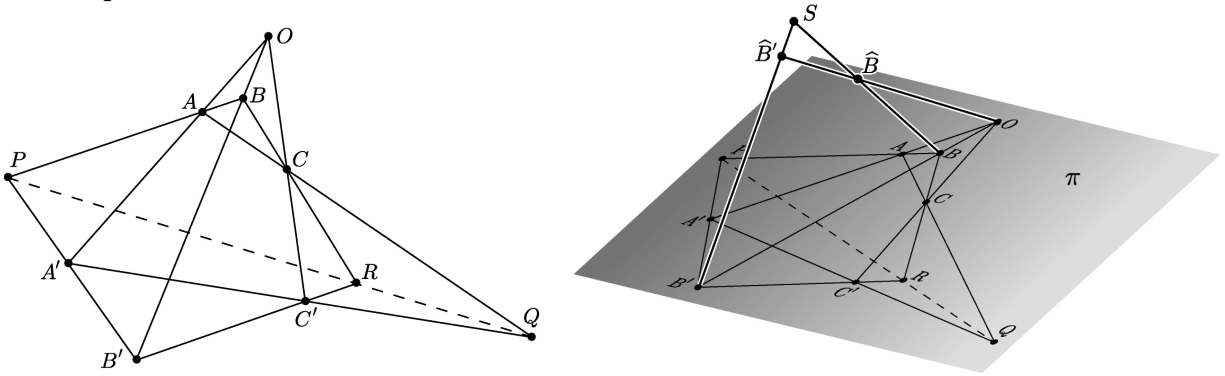
A projective 2-space is the same thing as a projective plane; and we have constructed non-classical examples of these. However for $n \geq 3$, every projective n-space is classical!

20.1 Theorem. Every projective n-space for $n \geq 3$ is isomorphic to $\mathbb{P}^n(K)$ for some skewfield K.

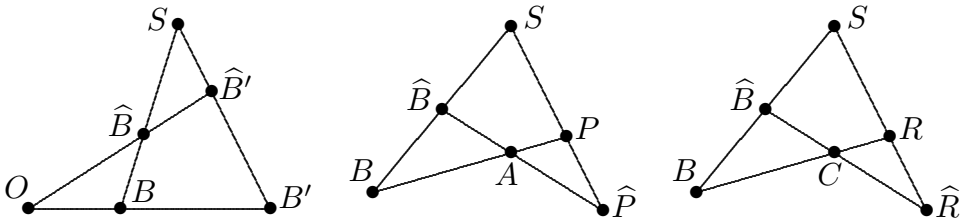
We present here just a few key ideas from the proof of Theorem 20.1. Let $(\mathfrak{P}, \mathfrak{L})$ be a projective n-space where $n \geq 3$. We first argue that

(20.2) Every plane of $(\mathfrak{P}, \mathfrak{L})$ is isomorphic to $\mathbb{P}^2(K)$ for some skewfield K.

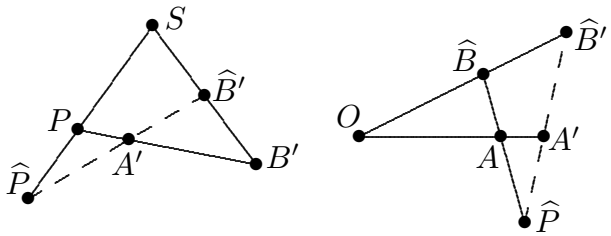
Let π be any plane in $(\mathfrak{P}, \mathfrak{L})$. Consider a point-line configuration in π as shown below; we must show that the three points P, Q, R are collinear. Let S be a point outside π, and let $\widehat{B}$ be a point on SB distinct from S and B.



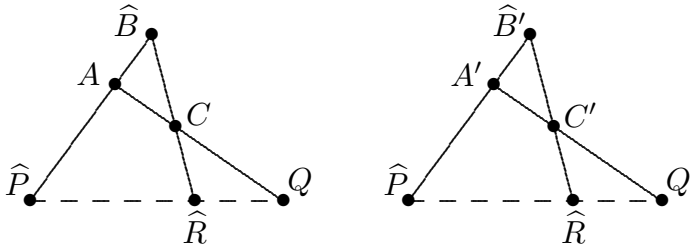
Let $\widehat{B}' = O\widehat{B} \cap SB'$, $\widehat{P} = A\widehat{B} \cap SP$, $\widehat{R} = \widehat{B}C \cap SR$; these points exist by the Veblen-Pasch axiom.



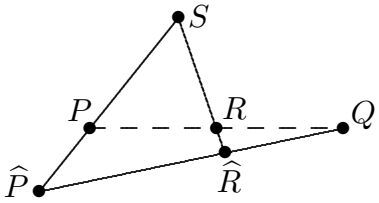
The points $\widehat{P}$, A' and $\widehat{B}'$ are collinear; this is because they lie in two distinct planes, as shown:



Now points $\widehat{P}, \widehat{R}, Q$ are also collinear since they lie in two distinct planes as shown:



Finally the points P, Q, R are collinear since they lie in two distinct planes—the original plane π and the plane shown:



By the converse of Desargues' Theorem (see comments at the end of Section 11) the plane π is isomorphic to $\mathbb{P}^2(K)$ for some skewfield K. This proves (20.2).

Next we prove that all planes in $(\mathfrak{P}, \mathfrak{L})$ are isomorphic to $\mathbb{P}^2(K)$ for the *same* skewfield K:

(20.3) Any two planes π, π' in $(\mathfrak{P}, \mathfrak{L})$ are isomorphic.

Consider first the case that $\pi' \cap \pi$ is a line; then π' and π lie in a common projective solid U (a projective 3-subspace). Let $P \in U$ be a point not in $\pi' \cup \pi$; then every line of U through P meets π in a unique point, and also meets π' in a unique point. This gives a bijection from points of π to points of π'. This bijection maps lines of π to lines of π': every plane β of U passing through P meets both π and π' in lines, which demonstrates the required correspondence. So $\pi \cong \pi'$, which proves (20.3) in the case $\pi' \cap \pi$ is a line.

Now in the general case it is easy to find a sequence of at most two intermediate planes π_1, π_2 of $(\mathfrak{P}, \mathfrak{L})$ such that

$$\pi \cong \pi_1 \cong \pi_2 \cong \pi';$$

here π shares a line with π_1, which shares a line with π_2, which shares a line with π'. For example if π and π' are disjoint (the worst case) take a flag (P, ℓ) in π and a flag (P', ℓ') in π'; then set $\pi_1 = \ell \vee P'$ and $\pi_2 = P \vee \ell'$. The other cases are similar, so (20.3) holds in general.

From here it is not hard to complete the proof that $(\mathfrak{P}, \mathfrak{L})$ itself is isomorphic to $\mathbb{P}^n(K)$.

Exercises 20.

1. Show that Axioms (S1)–(S6) above are not independent; that is, deduce one of these six statements from the other five.
2. Show that the Veblen-Pasch Axiom (V5) follows from (S1)–(S6), but is independent of (V1)–(V4).

21. Codes

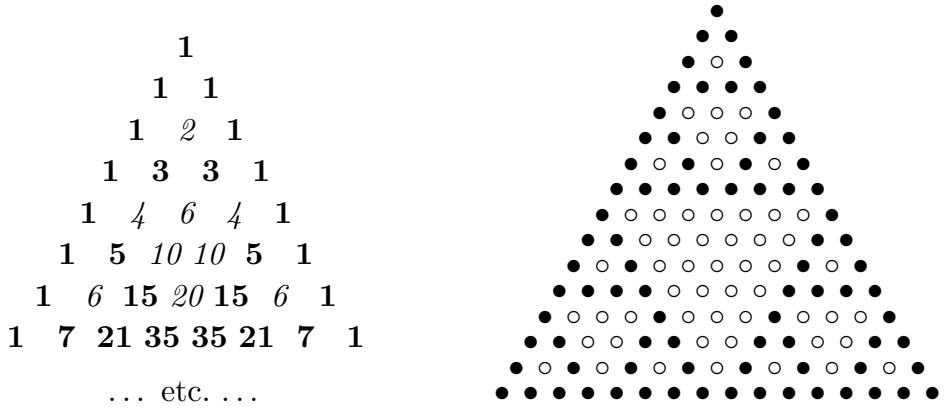
In Section 13 we found the p-rank of the incidence matrix of an arbitrary projective plane of order n, assuming $p \parallel n$, i.e. p divides n exactly once. More generally, however, the p-rank of a projective plane is not uniquely determined by its order, but rather depends

on the particular structure of the plane. A good rule-of-thumb is that among several nonisomorphic objects with the same parameters, the 'nicer' examples (in particular the more classical examples, or those with larger automorphism groups) tend to have lower p-ranks. In particular a nice formula is available for the p-rank of every classical projective plane of order $q = p^r$. We have waited until now to derive this formula since the same technique applies more generally to the incidence matrix of points versus hyperplanes of $\mathbb{P}^n(\mathbb{F}_q)$.

21.1 Theorem. Let A be the $N \times N$ incidence matrix of points versus hyperplanes of $\mathbb{P}^n(\mathbb{F}_q)$ where $q = p^r$ and $N = \begin{bmatrix} n+1 \\ 1 \end{bmatrix}_q = (q^{n+1}-1)/(q-1)$. Then the p-rank of A is

$$\binom{p+n-1}{n}^r + 1.$$

Check: By Theorem 13.3 the p-rank of $\mathbb{P}^2(\mathbb{F}_p)$ is $\binom{p+1}{2}+1$; and Theorem 21.1 confirms this. Our proof of Theorem 21.1 relies on a theorem of Lucas, which we introduce by looking at the pattern of *even* and **odd** numbers in Pascal's Triangle:



Most students will quickly recognize the emerging pattern as one that appears in *Sierpinski's Triangle.* The question of whether the binomial coefficient $\binom{m}{k}$ is even or odd is most readily answered by first expressing both m and k in binary as

$$m = m_0+2m_1+4m_2+8m_3+\cdots; \quad k = k_0+2k_1+4k_2+8k_3+\cdots \quad \text{where } m_i, k_i \in \{0,1\}.$$

Then

$$\binom{m}{k} \equiv \prod_{j\geq 0} \binom{m_j}{k_j} \bmod 2 \tag{21.2}$$

which vanishes whenever $k_j > m_j$ for some j, since $\binom{0}{1} = 0$. The product (21.2) is effectively finite since m and k have only finitely many nonzero binary digits, beyond which

all remaining factors are simply $\binom{0}{0} = 1$. Also if $k \notin \{0, 1, 2, \ldots, m\}$ then by definition $\binom{m}{k} = 0$; but then the right side of (21.2) also vanishes. The relation (21.2) generalizes to arbitrary primes, and also to multinomial coefficients. We recall the definition of multinomial coefficients:

$$\binom{m}{k_0,\ k_1,\ \cdots, k_n} = \begin{cases} \frac{m!}{k_0!k_1!\cdots k_n!}, & \text{if } m, k_0, \ldots, k_n \geq 0 \text{ and } k_0+k_1+\cdots+k_n = m; \\ 0, & \text{otherwise.} \end{cases}$$

21.3 Theorem (Lucas). Let p be prime, and express the non-negative integers $m, k_0, k_1, \ldots, k_n$ in base p as

$$m = \sum_{j\geq 0} m_j p^j; \qquad k_i = \sum_{j\geq 0} k_{ij} p^j$$

where $m_j, k_{ij} \in \{0, 1, 2, \ldots, p-1\}$. Then

$$\binom{m}{k_0,\ k_1,\ \cdots, k_n} \equiv \prod_{j\geq 0} \binom{m_j}{k_{0j},\ k_{1j},\ \cdots, k_{nj}} \mod p.$$

Proof. By the Multinomial Theorem,

$$(X_0 + X_1 + \cdots + X_n)^m = \sum_{k_0,\, k_1,\, \ldots,\, k_n} \binom{m}{k_0,\ k_1,\ \cdots,\ k_n} X_0^{k_0} X_1^{k_1} \cdots X_n^{k_n}. \tag{21.4}$$

There are only finitely many nonzero terms in the latter sum, namely those for which $k_0, k_1, \ldots, k_n \geq 0$ and $k_0+k_1+\cdots+k_n = m$. As a special case of (21.4), we first observe that

$$(X_0 + X_1 + \cdots + X_n)^p \equiv X_0^p + X_1^p + \cdots + X_n^p \mod p. \tag{21.5}$$

This follows from the fact that the multinomial coefficient

$$\binom{p}{k_0,\ k_1,\ \ldots,\ k_n} = \frac{p!}{k_0!k_1!\cdots k_n!}$$

is divisible by p unless one of the k_j's is p and all other k_j's are zero. Now (21.4) gives

$$\begin{aligned} (X_0 + X_1 + \cdots + X_n)^m &= \prod_{j\geq 0} (X_0 + X_1 + \cdots + X_n)^{p^j m_j} \\ &\equiv \prod_{j\geq 0} (X_0^{p^j} + X_1^{p^j} + \cdots + X_n^{p^j})^{m_j} \mod p \end{aligned}$$

by repeated application of (21.5), and so

$$\begin{aligned}
&(X_0{+}X_1{+}\cdots{+}X_n)^m \\
&\quad\equiv \prod_{j\ge 0} \sum_{k_{0j},\,k_{1j},\,\ldots,\,k_{nj}} \binom{m_j}{k_{0j},\,k_{1j},\,\cdots,\,k_{nj}} X_0^{k_{0j}p^j} X_1^{k_{1j}p^j} \cdots X_n^{k_{nj}p^j} \\
&\quad\equiv \sum_{\substack{k_{00},\,k_{01},\,\ldots;\\ k_{10},\,k_{11},\,\ldots;\\ \cdots\\ k_{n0},\,k_{n1},\,\ldots}} \Big[\prod_{j\ge 0} \tbinom{m_j}{k_{0j},k_{1j},\cdots,k_{nj}}\Big] X_0^{k_{00}+k_{01}p+k_{02}p^2+\cdots} X_1^{k_{10}+k_{11}p+k_{12}p^2+\cdots} \cdots X_n^{k_{n0}+k_{n1}p+k_{n2}p^2+\cdots} \\
&\qquad\qquad \mod p.
\end{aligned}$$

Here each $k_{ij} \in \{0,1,2,\ldots,p{-}1\}$ and so by the uniqueness of the base p expansion of the integers $m, k_0, \ldots, k_n$, we may equate coefficients with those in (21.4) to obtain the result. □

We are also interested in counting the number of terms in the multinomial expansion (21.4).

21.6 Lemma. The number of $(n{+}1)$-tuples $(k_0, k_1, \ldots, k_n)$ of non-negative integers such that $k_0{+}k_1{+}\cdots{+}k_n = m$ is

$$\binom{m+n}{n}.$$

Proof. The $(n{+}1)$-tuples $(k_0, k_1, \ldots, k_n)$ satisfying the required constraints, correspond bijectively to n-subsets of $\{0,1,2,\ldots,m{+}n{-}1\}$ via

$$(k_0, k_1, k_2, \ldots, k_n) \mapsto \big\{k_0,\ k_0{+}k_1{+}1,\ k_0{+}k_1{+}k_2{+}2,\ \ldots,\ k_0{+}k_1{+}\cdots{+}k_{n-1}{+}n{-}1\big\}. \quad \square$$

Proof of Theorem 21.1. Let A be the $N \times N$ incidence matrix of points versus hyperplanes of $\mathbb{P}^n(\mathbb{F}_q)$, where

$$N = \begin{bmatrix} n{+}1 \\ 1 \end{bmatrix}_q = 1{+}q{+}q^2{+}\cdots{+}q^n \equiv 1 \mod p.$$

By Exercise #19.3, every row and column of A has sum

$$\begin{bmatrix} n \\ 1 \end{bmatrix}_q = 1{+}q{+}q^2{+}\cdots{+}q^{n-1} \equiv 1 \mod p;$$

therefore the following matrix identity holds $\mod p$:

$$B^{\mathrm{T}} \left[\begin{array}{c|c} 0 & 0\cdots 0 \\ \hline 0 \\ \vdots & A \\ 0 \end{array}\right] B = \left[\begin{array}{c|c} 1 & 1\cdots 1 \\ \hline 1 \\ \vdots & A \\ 1 \end{array}\right] = B \left[\begin{array}{c|c} 1 & 0\cdots 0 \\ \hline 0 \\ \vdots & A{-}J \\ 0 \end{array}\right] B^{\mathrm{T}}$$

where the matrices

$$B = \left[\begin{array}{c|c} 1 & 0\cdots 0 \\ \hline 1 & \\ \vdots & I_N \\ 1 & \end{array}\right] \quad \text{and} \quad J = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

have size $(N{+}1){\times}(N{+}1)$ and $N{\times}N$ respectively, and I_N denotes the $N{\times}N$ identity matrix. This yields

$$\operatorname{rank}_p A = 1 + \operatorname{rank}_p(A{-}J) = 1 + \operatorname{rank}_p(J{-}A).$$

The matrix $J{-}A$ is easier to describe algebraically than A, as we now observe. Every nonzero vector $a = (a_0, a_1, \ldots, a_n) \in \mathbb{F}_q^{n+1}$ determines a hyperplane H_a consisting of all points $\langle x\rangle{=}\langle(x_0, x_1, \ldots, x_n)\rangle \le \mathbb{F}_q^{n+1}$ such that

$$a_0x_0 + a_1x_1 + \cdots + a_nx_n = 0.$$

Note that $H_{\lambda a} = H_a$ whenever $\lambda \neq 0$; therefore the hyperplane H_a is determined by the 1-space $\langle a\rangle \le \mathbb{F}_q^{n+1}$. The $(\langle x\rangle, \langle a\rangle)$-entry of $J{-}A$ is

$$(a_0x_0 + a_1x_1 + \cdots + a_nx_n)^{q-1} = \begin{cases} 0, & \text{if the point } \langle x\rangle \text{ lies in } H_a; \\ 1, & \text{otherwise.} \end{cases}$$

We consider therefore the $q^{n+1}{\times}q^{n+1}$ matrix M with rows and columns indexed by vectors $x, a \in \mathbb{F}_q^{n+1}$, and whose (x, a)-entry is $(a_0x_0 + a_1x_1 + \cdots + a_nx_n)^{q-1}$. This matrix is larger than $J{-}A$ itself; in particular it has an extra row and column of zeroes, corresponding to $x{=}0$ and $a{=}0$. It also has many repeated rows and columns, since for all $\lambda \neq 0$ the rows of M indexed by x and λx coincide; also the columns of M indexed by a and λa coincide. Evidently $J{-}A$ is obtained from M by deleting duplicate rows and columns, as well as deleting a row of zeroes and a column of zeroes, so that

$$\operatorname{rank}_p A = 1 + \operatorname{rank}_p(J{-}A) = 1 + \operatorname{rank}_p M. \tag{21.7}$$

The matrix M can be conveniently expressed in terms of another $q^{n+1}{\times}q^{n+1}$ matrix U with rows indexed by vectors $x \in \mathbb{F}_q^{n+1}$ and columns indexed by $(n{+}1)$-tuples $k = (k_0, k_1, \ldots, k_n) \in \{0, 1, 2, \ldots, q{-}1\}^{n+1}$; the (x, k)-entry of U is

$$u_{x,k} = x_0^{k_0} x_1^{k_1} \cdots x_n^{k_n}.$$

Also define the $q^{n+1}{\times}q^{n+1}$ diagonal matrix D with rows and columns indexed by $(n{+}1)$-tuples $k, \ell \in \{0, 1, 2, \ldots, q{-}1\}^{n+1}$; the (k, ℓ)-entry of D is

$$d_{k,\ell} = \binom{q-1}{k_0,\ k_1,\ \ldots,\ k_n} \delta_{k,\ell}$$

where we use the Kronecker delta symbol

$$\delta_{k,\ell} = \begin{cases} 1, & \text{if } k = \ell, \text{ i.e. } (k_0, k_1, \ldots, k_n) = (\ell_0, \ell_1, \ldots, \ell_n); \\ 0, & \text{otherwise.} \end{cases}$$

Now the (x, a)-entry of UDU^{T} is

$$\begin{aligned} \sum_{k,\ell} u_{x,k} d_{k,\ell} u_{a,\ell} &= \sum_{k} d_{k,k} u_{a,k} u_{x,k} \\ &= \sum_{k_0,\, k_1,\, \ldots,\, k_n} \binom{q-1}{k_0,\ k_1,\ \ldots,\ k_n} a_0^{k_0} a_1^{k_1} \cdots a_n^{k_n} x_0^{k_0} x_1^{k_1} \cdots x_n^{k_n} \\ &= (a_0 x_0 + a_1 x_1 + \cdots + a_n x_n)^{q-1} \end{aligned}$$

and so

$$M = UDU^{\mathrm{T}}. \tag{21.8}$$

We now observe that

(21.9) the square matrix U is nonsingular.

For suppose a vector $v = \left(c_k \,:\, k \in \{0, 1, 2, \ldots, q-1\}^{n+1}\right)$ satisfies $Uv^{\mathrm{T}} = 0$; then

$$\sum_{k_0,\, k_1,\, \ldots,\, k_n} c_{k_0, k_1, \ldots, k_n} x_0^{k_0} x_1^{k_1} \cdots x_n^{k_n} = 0$$

for all $x = (x_0, x_1, \ldots, x_n) \in \mathbb{F}_q^{n+1}$. This forces all coefficients to vanish, i.e. $c_k = 0$, which proves (21.9).

Combining (21.7) through (21.9), the p-rank of A equals the number of $(n{+}1)$-tuples $k = (k_0, k_1, \ldots, k_n)$ such that the multinomial coefficient $\binom{q-1}{k_0,\, k_1,\, \ldots,\, k_n}$ is not divisible by p. The answer is found using Lucas' Theorem: Consider the base p representations

$$\begin{aligned} q{-}1 &= (p{-}1) + (p{-}1)p + (p{-}1)p^2 + \cdots + (p{-}1)p^{r-1}; \\ k_i &= k_{i0} {+} k_{i1} p {+} k_{i2} p^2 {+} \cdots {+} k_{i,r-1} p^{r-1}, \quad k_{ij} \in \{0, 1, 2, \ldots, p{-}1\}. \end{aligned}$$

We have

$$\begin{aligned} p \nmid \binom{q-1}{k_0,\, k_1,\, \ldots,\, k_n} \quad &\text{iff} \quad p \nmid \prod_{0 \le j < r} \binom{p-1}{k_{0j},\, k_{1j},\, \ldots,\, k_{nj}} \\ &\text{iff} \quad \text{for all } j \in \{0, 1, 2, \ldots, r{-}1\} \text{ we have } k_{0j} {+} k_{1j} {+} \cdots {+} k_{nj} = p{-}1. \end{aligned}$$

By Theorem 21.6, for each j there are $\binom{p-1+n}{n}$ choices of $(n{+}1)$-tuple $(k_{0j}, k_{1j}, \ldots, k_{nj}) \in \{0, 1, 2, \ldots, p{-}1\}^{n+1}$ such that $k_{0j} {+} k_{1j} {+} \cdots {+} k_{nj} = p{-}1$; this gives exactly $\binom{p-1+n}{n}^r$

choices of $(k_0, k_1, \ldots, k_n) \in \{0, 1, 2, \ldots, q-1\}^{n+1}$ such that $p \nmid \binom{q-1}{k_0,\, k_1,\, \ldots,\, k_n}$. The conclusion of Theorem 21.1 follows. □

Exercises 21.

1. A **frame** in $\mathbb{P}^n(F)$ is a set of $n+2$ points, no $n+1$ of which lie in a hyperplane. (Thus for example a frame in $\mathbb{P}^2(F)$ is a quadrangle: 4 points with no three collinear; see also Theorem 19.4.) If A is a point-hyperplane incidence matrix for $\mathbb{P}^n(\mathbb{F}_2)$, show that the rows of A corresponding to a frame, form a basis for the row space of A over $\mathbb{F}_2$.

   Hint. Use Theorem 21.1.

22. The Plücker Map

The Plücker map introduces coordinates for the k-subspaces of a finite-dimensional vector space over an arbitrary field F. We introduce this topic by discussing the coordinatization of lines in $\mathbb{P}^3(F)$. This important special case leads naturally to the Klein correspondence (Section 25).

22.1 Lemma. Let F be an arbitrary field. For every 2-subspace $U < F^4$ there is a unique (up to nonzero scalar multiple) skew-symmetric 4×4 matrix over F with row space equal to U. Such a matrix is given by $u^{\mathrm{T}}v - v^{\mathrm{T}}u$ where $\{u, v\}$ is a basis for U.

Proof. Let $U = \{u, v\} < F^4$ be a 2-subspace. There exists $\gamma \in F^4$ such that $\gamma u^{\mathrm{T}} = 0$ and $\gamma v^{\mathrm{T}} = 1$, so $\gamma(u^{\mathrm{T}}v - v^{\mathrm{T}}u) = u$ lies in the row space of $u^{\mathrm{T}}v - v^{\mathrm{T}}u$; similarly v lies in its row space. Thus the row space of $u^{\mathrm{T}}v - v^{\mathrm{T}}u$ contains $U = \langle u, v\rangle$ and so must equal U.

If $\{\widetilde{u}, \widetilde{v}\}$ is also basis for U then there exists a unique $A \in GL_2(F)$ such that $\widetilde{u}, \widetilde{v}$ are the rows of LA, where L is the 2×4 matrix with rows u, v. Now

$$\widetilde{u}^{\mathrm{T}}\widetilde{v} - \widetilde{v}^{\mathrm{T}}\widetilde{u} = LA \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} A^{\mathrm{T}}L^{\mathrm{T}} = (\det A)L \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} L^{\mathrm{T}} = (\det A)\big(u^{\mathrm{T}}v - v^{\mathrm{T}}u\big)$$

so the matrix $u^{\mathrm{T}}v - v^{\mathrm{T}}u$ changes only by a nonzero scalar multiple.

Finally let M be any skew-symmetric 4×4 matrix whose row space is U; we must show that $kM = x^{\mathrm{T}}y - y^{\mathrm{T}}x$ for some nonzero $k \in F$ and some basis $\{x, y\}$ of U. We may suppose

$$M = \begin{bmatrix} 0 & 1 & a & b \\ -1 & 0 & c & d \\ -a & -c & 0 & e \\ -b & -d & -e & 0 \end{bmatrix}$$

for some $a, b, c, d, e \in F$ since some entry of M is nonzero, and we may multiply M by a nonzero scalar to make this entry 1. (The argument is similar if the nonzero entry of M is in any other position.) Since the rank of M is 2, we have

$$0 = \det \begin{bmatrix} 0 & 1 & b \\ -1 & 0 & d \\ -a & -c & e \end{bmatrix} = e - ad + bc$$

and so $M = x^{\mathrm{T}}y - y^{\mathrm{T}}x$ where x, y are the first two rows of M. □

If $\ell = \langle u, v \rangle$ is any line (i.e. vector 2-space) of F^4 then the corresponding 4×4 matrix as in Lemma 22.1 has the form

$$(\ell_{01}, \ell_{02}, \ell_{03}, \ell_{12}, \ell_{13}, \ell_{23}) := \begin{bmatrix} 0 & \ell_{01} & \ell_{02} & \ell_{03} \\ -\ell_{01} & 0 & \ell_{12} & \ell_{13} \\ -\ell_{02} & -\ell_{12} & 0 & \ell_{23} \\ -\ell_{03} & -\ell_{13} & -\ell_{23} & 0 \end{bmatrix} = u^{\mathrm{T}}v - v^{\mathrm{T}}u$$

where we abbreviate an arbitrary skew-symmetric 4×4 matrix by the 6-tuple of its above-diagonal entries; here

$$\ell_{ij} = -\ell_{ji} = \det \begin{bmatrix} u_i & u_j \\ v_i & v_j \end{bmatrix} = u_i v_j - u_j v_i \,.$$

We refer to $(\ell_{01}, \ell_{02}, \ell_{03}, \ell_{12}, \ell_{13}, \ell_{23})$ as the **Plücker coordinates** of the line ℓ. Denote by S the vector 6-space of all skew-symmetric 4×4 matrices over F. By Exercise #1, every nonzero element of S has rank 2 or 4; accordingly we distinguish the points of S as having either rank 2 or rank 4. The **Plücker map** is the bijection

$$\begin{aligned} \rho \;:\; \{\text{lines of } \mathbb{P}^3(F)\} &\longrightarrow \{\text{'rank 2' points of } \mathbb{P}S\} \\ \ell = \langle u, v \rangle &\mapsto \langle (\ell_{01}, \ell_{02}, \ell_{03}, \ell_{12}, \ell_{13}, \ell_{23}) \rangle = \langle u^{\mathrm{T}}v - v^{\mathrm{T}}u \rangle \end{aligned}$$

using our abbreviation of elements of S as 6-tuples.

22.2 Theorem. (i) The Plücker map ρ is a bijection from the set of lines in $\mathbb{P}^3(F)$ and the set of 'rank 2' points of $\mathbb{P}S = \mathbb{P}^5(F)$.

(ii) The Plücker coordinates of ℓ are also determined (to within a nonzero scalar multiple) by any two distinct planes containing ℓ (as above) by the formula

$$\ell_{ij} = -\ell_{ji} = a_k b_m - a_m b_k$$

where (i, j, k, m) is an even permutation of $(0, 1, 2, 3)$.

(iii) A point $\langle z \rangle = \langle (z_0, z_1, z_2, z_3) \rangle$ lies in ℓ iff

$$\begin{bmatrix} 0 & \ell_{23} & -\ell_{13} & \ell_{12} \\ -\ell_{23} & 0 & \ell_{03} & -\ell_{02} \\ \ell_{13} & -\ell_{03} & 0 & \ell_{01} \\ -\ell_{12} & \ell_{02} & -\ell_{01} & 0 \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ z_3 \end{bmatrix} = 0.$$

(iv) A plane $\gamma^\perp = (c_0, c_1, c_2, c_3)^\perp$ contains ℓ iff

$$\begin{bmatrix} 0 & \ell_{01} & \ell_{02} & \ell_{03} \\ -\ell_{01} & 0 & \ell_{12} & \ell_{13} \\ -\ell_{02} & -\ell_{12} & 0 & \ell_{23} \\ -\ell_{03} & -\ell_{13} & -\ell_{23} & 0 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = 0.$$

(v) The Plücker coordinates satisfy $\ell_{01}\ell_{23} - \ell_{02}\ell_{13} + \ell_{03}\ell_{12} = 0$.

Proof. Conclusion (i) has already been verified. Consider now an arbitrary line $\ell = \langle x, y \rangle$ as above. A point $\langle z \rangle = \langle (z_0, z_1, z_2, z_3) \rangle$ lies on ℓ iff the matrix

$$\begin{bmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \\ z_0 & z_1 & z_2 & z_3 \end{bmatrix}$$

has rank 2, which is to say that all four 3×3 submatrices (formed by deleting one column at a time) are singular; for example this implies that

$$0 = \det \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} = \ell_{23} z_1 - \ell_{13} z_2 + \ell_{12} z_3 \,.$$

By similarly expanding determinants for all 3×3 submatrices, we obtain (iii). Also since the columns of $x^{\mathrm{T}} y - y^{\mathrm{T}} x$ lie in ℓ, they must satisfy the condition (iii); this yields (v).

In view of (iii) we are led to consider the linear transformation $R : S \to S$ defined by

$$R : \begin{bmatrix} 0 & c_{01} & c_{02} & c_{03} \\ -c_{01} & 0 & c_{12} & c_{13} \\ -c_{02} & -c_{12} & 0 & c_{23} \\ -c_{03} & -c_{13} & -c_{23} & 0 \end{bmatrix} \mapsto \begin{bmatrix} 0 & c_{23} & -c_{13} & c_{12} \\ -c_{23} & 0 & c_{03} & -c_{02} \\ c_{13} & -c_{03} & 0 & c_{01} \\ -c_{12} & c_{02} & -c_{01} & 0 \end{bmatrix}.$$

Thus for every line $\ell = \langle x, y\rangle$ we have

$$\ell = \ker R(x^{\mathrm{T}}y - y^{\mathrm{T}}x) = \text{row space of } x^{\mathrm{T}}y - y^{\mathrm{T}}x. \tag{22.3}$$

Here we may identify row and column space via the transpose map, since $x^{\mathrm{T}}y - y^{\mathrm{T}}x$ is skew-symmetric; similarly

$$\ker R(x^{\mathrm{T}}y - y^{\mathrm{T}}x) = \text{left null space of } R(x^{\mathrm{T}}y - y^{\mathrm{T}}x)$$

is identified with the right null space via the transpose map. A plane $\gamma^\perp$ contains ℓ iff $\gamma \in \ell^\perp$ iff $(x^{\mathrm{T}}y - y^{\mathrm{T}}x)\gamma^{\mathrm{T}} = 0$, which yields (iv).

Assuming $\ell = \alpha^\perp \cap \beta^\perp = \langle \alpha, \beta\rangle^\perp$ where $\alpha, \beta \in F^4$ are linearly independent, from (22.3) we have

$$\langle \alpha, \beta\rangle = \ker R(\alpha^{\mathrm{T}}\beta - \beta^{\mathrm{T}}\alpha) = \text{row space of } \alpha^{\mathrm{T}}\beta - \beta^{\mathrm{T}}\alpha$$

and so

$$\ell = \langle \alpha, \beta\rangle^\perp = \ker(\alpha^{\mathrm{T}}\beta - \beta^{\mathrm{T}}\alpha) = \text{row space of } R(\alpha^{\mathrm{T}}\beta - \beta^{\mathrm{T}}\alpha).$$

Since S has a unique member (up to nonzero scalar multiple) with row space equal to ℓ, we have

$$R(\alpha^{\mathrm{T}}\beta - \beta^{\mathrm{T}}\alpha) = k(x^{\mathrm{T}}y - y^{\mathrm{T}}x)$$

for some nonzero $k \in F$. This yields (ii). □

As promised, we now describe the more general Plücker map which coordinatizes vector k-subspaces of an n-space; that is, projective $(k-1)$-subspaces of $\mathbb{P}^{n-1}(F)$. Associated to a k-subspace $U \leq F^n$ one obtains an $\binom{n}{k}$-tuple of coordinates $u_{i_1 i_2 \cdots i_k} \in F$ for all subscripts satisfying $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. The resulting vector of Plücker coordinates, which may be computed from any choice of basis for U, is well-defined (i.e. independent of the choice of basis) up to nonzero scalar multiple. This vector determines a point $\rho(U)$ in $\mathbb{P}^{N-1}(F)$ where $N = \binom{n}{k}$; here ρ is the Plücker map. The image of this map in general is a (typically small) subset of the points of $\mathbb{P}^{N-1}(F)$ known as a **Grassmannian** or **Grassmann variety** (or rather the set of F-rational points of this variety, if F is not algebraically closed). In the case $k = 2$ and $n = 4$ the Plücker map associates to every line (vector 2-subspace) $\ell < F^4$ the point

$$\rho(\ell) = \langle (\ell_{ij})_{0 \leq i < j \leq 3}\rangle$$

in $\mathbb{P}^5(F)$. In this case the Grassmann variety is the point set in $\mathbb{P}^5(F)$ defined by the single equation $\ell_{01}\ell_{23} - \ell_{02}\ell_{13} + \ell_{03}\ell_{12} = 0$, which is a quadric. More general Grassmann varieties (for larger values of n and k) are not quadrics, but can be described as intersections of quadrics.

The Plücker map in the full generality we are describing here is most conveniently described in the language of *exterior algebra* (see Appendix A4): one simply maps a vector k-subspace $U \leq F^n$ to the vector 1-subspace $\bigwedge^k U \leq \bigwedge^k F^n$. This gives a well-defined map from the projective $(k-1)$-subspace $\mathbb{P}U$ to the point $\rho(U) = \mathbb{P}(\bigwedge^k U)$ in $\mathbb{P}^k(\bigwedge^k F) = \mathbb{P}^{N-1}(F)$ where $N = \binom{n}{k}$. If one specifies U by choosing a $k \times n$ matrix A whose row space is U, then the coordinates of $\rho(U)$ are the $k \times k$ minors of A; there are N of these. Although the choice of matrix A is not uniquely determined by U, different choices of A will give the same N-tuple of coordinates of $\rho(U)$, up to nonzero scalar multiple. This generalizes the case for $k = 2$ and $n = 4$ described above, where the Plücker coordinates of $\ell = \langle u, v \rangle$ are just the 2×2 minors of the matrix $\left[\begin{smallmatrix} u_1 & u_2 & u_3 & u_4 \\ v_1 & v_2 & v_3 & v_4 \end{smallmatrix}\right]$.

Exercises 22.

1. Let F be an arbitrary field. Show that every skew-symmetric matrix over F has the form BAB^{T} where B is invertible and A is block-diagonal with blocks of the form either $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ or $[0]$. Conclude that every skew-symmetric matrix has even rank.

   Hint. Let M be a skew-symmetric matrix over F. If A is nonzero, show that a sequence of elementary operations can be performed simultaneously on the rows and columns of M, transforming it to a block-diagonal matrix with upper-left corner $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$. Apply induction.

2. Show that for every skew-symmetric 4×4 matrix $M = (c_{ij} : 0 \leq i, j \leq 3)$, we have $\det M = (c_{01}c_{23} - c_{02}c_{13} + c_{03}c_{12})^2$. Conclude that the image of the Plücker map ρ of Theorem 22.2 consists of all points of $\mathbb{P}S$ satisfying the necessary condition of conclusion (v).

   Hint. Use Exercise #1.

23. Quadratic Forms

After studying subspaces of a projective space (these being the solutions of linear systems, hence varieties of degree 1) the next natural objects of study are quadrics (varieties of degree 2), these being the zeroes of quadratic forms. For example, quadrics in the classical projective planes are simply conics, which we have previously considered. Insight into the geometry of more general quadrics is founded upon a solid understanding of quadratic forms. In this section we lay the algebraic foundations for quadratic forms; and in the next section we study more geometric questions concerning the associated quadrics. Some of this terminology was previously introduced in Section 14; here it is presented in more generality.

Let $V = F^n$ where the field F is for the moment arbitrary. A **quadratic form** on V is a map $Q : V \to F$ defined by a homogeneous polynomial of degree 2 in the coordinates. Equivalently, for all $x, y \in V$ and $c \in F$ we have

(Q1) $Q(cx) = c^2 Q(x)$;

(Q2) $Q(x + y) = Q(x) + Q(y) + B(x, y)$

where $B(x, y)$ is a symmetric bilinear form. Note that the bilinear form B is uniquely determined by the quadratic form Q using the *polarization identity* (Q2). Every quadratic form, and its associated bilinear form, may be written as

$$Q(x) = xAx^{\mathrm{T}}; \qquad B(x, y) = x(A + A^{\mathrm{T}})y$$

for some $n \times n$ matrix $A = \big(a_{ij}\big)$ over F.

Let V and W be vector spaces over a field F, and let $Q : V \to F$ and $Q' : W \to F$ be quadratic forms. An **isometry** from (V, Q) to (V', Q') (or from Q to Q') is a vector space isomorphism $\theta : V \to V'$ such that $Q'(\theta(v)) = Q(v)$ for all $v \in V$. (Strictly speaking, such a map θ is a *linear isometry;* more general isometries are invertible *semilinear* transformations preserving the quadratic form.) The relation of isometry is the most natural equivalence relation on the set of quadratic forms. The isometries from (V, Q) to itself form a group, called the **orthogonal group** $O(V, Q)$. By virtue of (Q2) we see that every isometry $\theta \in O(V, Q)$ is also an **isometry** of B, i.e. $B(\theta(x), \theta(y)) = B(x, y)$ for all $x, y \in V$. The converse fails in characteristic 2 as we observe below.

A typical vector space isomorphism $\theta : V \to V$ has the form $x \mapsto xM$ where $M \in GL_n(F)$. For $Q(x) = xAx^{\mathrm{T}}$ we have

$$Q(\theta(x)) = Q(xM) = xMAM^{\mathrm{T}}x^{\mathrm{T}}$$

so that θ is an isometry from Q to the quadratic form Q' defined by the matrix MAM^{T}. We say that two $n \times n$ matrices A, B are **congruent** if $B = MAM^{\mathrm{T}}$ for some $M \in GL_n(F)$. This is an equivalence relation on the set of all $n \times n$ matrices over F, and congruent matrices represent isometric quadratic forms.

For each subspace $U \leq V$ we define $U^{\perp} = \{v \in V : B(u, v) = 0 \text{ for all } u \in U\}$. The **radical** of U is $U^{\perp} \cap U$; in particular the radical of V is simply $V^{\perp}$. We say that the bilinear form B is **nondegenerate** if $V^{\perp} = 0$; in this case elementary linear algebra shows that $\dim U + \dim U^{\perp} = \dim V$ for all $U \leq V$. In general however the subspace $U^{\perp}$ need not be disjoint from U; the subspaces U and $U^{\perp}$ may overlap considerably, and even coincide. Note that B is nondegenerate iff the matrix $A + A^{\mathrm{T}}$ is nonsingular. A point $\langle u \rangle$ is called **singular** if $Q(u) = 0$; or **nonsingular** if $Q(u) \neq 0$. A subspace $U \leq V$ is **totally singular** if $Q(u) = 0$ for all $u \in U$. We say that a subspace U is **totally isotropic** if $B(u, u') = 0$ for all $u, u' \in U$, i.e. $U \leq U^{\perp}$. Every totally singular subspace is totally isotropic; the converse holds only in characteristic $\neq 2$. We say Q is **degenerate** if the radical $V^{\perp}$ contains a singular point; otherwise Q is **nondegenerate**. Thus if B is nondegenerate, so is Q; the converse holds in odd characteristic. If Q is nondegenerate then by preceding remarks, every totally singular subspace has dimension at most $\lfloor \frac{1}{2} \dim V \rfloor$. A **maximal totally singular subspace** is a totally singular subspace U which is not contained in any larger totally singular subspace.

Assume V is a vector space with basis $\{v_1, v_2, \ldots, v_n\}$ over a field F. The **discriminant** of a quadratic form $Q : V \to F$, denoted $\operatorname{disc} Q$, is the determinant of the **Gram matrix** $\big(B(v_i, v_j) : 1 \leq i, j \leq n\big)$. Assuming $Q(x) = xAx^{\mathrm{T}}$, we have

$$\operatorname{disc} Q = \det(A + A^{\mathrm{T}}).$$

The discriminant of Q is defined *only up to multiplication by a nonzero square in F,* since a change of basis for V (using a change-of-basis matrix M, say) replaces A by MAM^{T}, and we obtain

$$\det(M(A + A^{\mathrm{T}})M^{\mathrm{T}}) = (\det M)^2 \det(A + A^{\mathrm{T}}).$$

With this understanding, the discriminant depends only on the isometry type of Q. Note that if F is a finite field of characteristic 2 then every element of F is a square so the discriminant contains no useful information. However if $F = \mathbb{F}_q$ where q is odd then half the nonzero elements of F are squares and the other half are nonsquares; so assuming Q is nondegenerate, the value of $\operatorname{disc} Q$ contains one bit of information, i.e. 'square' or 'nonsquare'.

Let $V_1, V_2, \ldots, V_r$ be vector spaces over F with corresponding quadratic forms $Q_i : V_i \to F$. The **orthogonal direct sum** of the V_i's is the vector space $V = V_1 \oplus V_2 \oplus \cdots \oplus V_r$ equipped with the quadratic form

$$Q(v_1 + v_2 + \cdots + v_r) = Q_1(v_1) + Q_2(v_2) + \cdots + Q_r(v_r) \quad \text{for } v_i \in V_i\,.$$

Such an *orthogonal* direct sum is often written $V = V_1 \perp V_2 \perp \cdots \perp V_r$. If Q_i is defined by a square matrix A_i, then Q is defined by the block-diagonal matrix $A_1 \oplus A_2 \oplus \cdots \oplus A_r$ and

$$\operatorname{disc} Q \;=\; \prod_{1\le i\le r} \operatorname{disc} Q_i \;=\; \prod_{1\le i\le r} \det(A_i + A_i^{\mathrm{T}})\,.$$

23.1 Real Quadratic Forms. Consider a quadratic form $Q : \mathbb{R}^n \to \mathbb{R}$ given by $Q(x) = xAx^{\mathrm{T}}$ where A is a real $n\times n$ matrix. We may suppose A is symmetric; otherwise replace A by its symmetric part $\frac{1}{2}(A + A^{\mathrm{T}})$ without changing the value of $Q(x)$ (because $xA^{\mathrm{T}}x^{\mathrm{T}} = xAx^{\mathrm{T}}$). Recall that every real symmetric matrix is diagonalizable by an orthogonal change of basis, so that $A = MDM^{\mathrm{T}}$ where $M^{\mathrm{T}} = M^{-1}$ and $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Now

$$Q(xM) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2\,.$$

By a further change of basis we may assume all $\lambda_i \in \{-1, 0, 1\}$; for if $\lambda_i \neq 0$ we replace $x_i \mapsto |\lambda_i|^{-1/2} x_i$. Finally we may permute coordinates to see that $Q(x)$ is isometric to

$$x_1^2 + \cdots + x_k^2 - x_{k+1}^2 - \cdots - x_m^2$$

where $0 \le k \le m \le n$. The number of isometry classes of quadratic forms on $\mathbb{R}^n$ is the number of pairs of integers (k, m) satisfying these constraints, namely $\binom{n+2}{2} = \frac{1}{2}(n+1)(n+2)$. The form Q is nondegenerate iff A is nonsingular, iff $m{=}n$. There are $n{+}1$ isometry classes of nondegenerate quadratic forms on $\mathbb{R}^n$, ranging from positive definite ($k{=}m{=}n$) to negative definite ($k{=}0$, $m{=}n$). In the nondegenerate case $Q(x) = x_1^2 + \cdots + x_k^2 - x_{k+1}^2 - \cdots - x_n^2$ has maximal totally singular subspaces of dimension $\min\{k, n{-}k\} \le \lfloor \frac{n}{2} \rfloor$; an example of such a subspace is the subspace spanned by the vectors $e_i + e_{k+1}$ for $i = 1, 2, \ldots, \min\{k, n{-}k\}$, where $\{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{R}^n$.

23.2 Odd Characteristic. Suppose F has characteristic different from 2 (possibly characteristic zero). We will assume (as we may, for the same reasons as in the real case) that

A is symmetric. We may recover the quadratic form Q from the associated bilinear form since

$$B(x,x) = Q(2x) - 2Q(x) = 4Q(x) - 2Q(x) = 2Q(x)$$

from which we obtain

$$Q(x) = \tfrac{1}{2}B(x,x).$$

From this identity it follows that isometries of the bilinear form B are the same as isometries of the quadratic form Q: if $\theta \in GL(V)$ satisfies $B(\theta(x),\theta(y)) = B(x,y)$ for all $x, y \in V$ then $Q(\theta(x)) = \frac{1}{2}B(\theta(x),\theta(x)) = \frac{1}{2}B(x,x) = Q(x)$ and the converse follows from (Q2). Also in this case a subspace $U \leq V$ is totally isotropic iff it is totally singular. Moreover Q is nondegenerate iff B is, iff A is nonsingular.

23.3 Even Characteristic. Suppose F has characteristic 2. In this case we cannot assume A is symmetric; when convenient, however, we may assume A is upper triangular. We obtain $B(x,x) = 2Q(x) = 0$ and so the associated bilinear form is **alternating**; in other words the matrix $A+A^{\mathrm{T}}$ associated to B is skew-symmetric with zero diagonal. By Exercise #22.1, the matrix $A+A^{\mathrm{T}}$ has even rank, so it cannot be nonsingular unless n is odd. We cannot recover the quadratic form from the bilinear form. For example when $n = 2$, the two forms $Q_1(x_1,x_2,x_3) = x_1^2 + x_2x_3$ and $Q_2(x_1,x_2,x_3) = x_2x_3$ give rise to the same degenerate bilinear form $B(x,y) = x_2y_3 + x_3y_2$. Both have radical $\langle(1,0,0)\rangle$ but in the case of Q_1 this radical is nonsingular, whereas in the case of Q_2 the radical is singular. Thus Q_1 is nondegenerate whereas Q_2 is degenerate. Note that the quadric defined by Q_2 (i.e. the conic defined by the zeroes of Q_2) is a pair of intersecting lines $\langle(0,1,0)^{\perp}\rangle$ and $\langle(0,0,1)^{\perp}\rangle$, a clearly degenerate conic; the point $\langle(1,0,0)\rangle$ of intersection is the singular radical. On the other hand Q_1 defines a nondegenerate conic; in this case the nonsingular radical $\langle(1,0,0)\rangle$ is the nucleus of the conic.

Most questions about general quadratic forms reduce to questions about the nondegenerate case. This reduction follows from the following observation.

23.4 Lemma. Let V be a vector space over a field F, and let $Q : V \to F$ be a degenerate quadratic form whose radical $V^{\perp}$ contains a totally singular subspace U. Then Q naturally induces a quadratic form on V/U.

Proof. We naturally define $\widetilde{Q} : V/U \to F$ by $v + U \mapsto Q(v)$. Note that this is well-defined: for all $v \in V$ and $u \in U$,

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

One easily checks that $\widetilde{Q}$ satisfies the properties (Q1) and (Q2), using the similar properties for Q. □

Classification of quadratic forms proceeds by induction using the following lemma.

23.5 Lemma. Let V be a vector space over a field F of finite dimension n, and let $Q : V \to F$ be a quadratic form. Suppose $\langle x\rangle$ is a singular point not contained in the radical $V^\perp$. Then

(a) There exists a singular point $\langle y\rangle$ such that $V = \langle x, y\rangle \oplus \langle x, y\rangle^\perp$ and
$$Q(ax + by + z) = ab + Q(z)$$
for all $a, b \in F$ and $z \in \langle x, y\rangle^\perp$.

(b) The radical of $\langle x, y\rangle^\perp$ equals $V^\perp$. In particular Q is nondegenerate iff the restriction of Q to the $(n-2)$-space $\langle x, y\rangle^\perp$ is nondegenerate.

A pair of vectors $\{x, y\}$ such that $Q(x) = Q(y) = 0$ and $B(x, y) = 1$, as above, is a **hyperbolic pair**.

Proof of Lemma 23.5. Since $x \notin V^\perp$, there exists $y \in V$ such that $B(x, y) \neq 0$. For all $c \in F$,
$$Q(cx + y) = c^2Q(x) + Q(y) + cB(x, y) = Q(y) + cB(x, y),$$
and since $B(x, y) \neq 0$, there is a unique $c \in F$ such that $Q(cx + y) = 0$. Set $y' = cx+y$. Now $\langle x\rangle$ and $\langle y'\rangle$ are singular points with $B(x, y') \neq 0$. Set $y'' = B(x, y')^{-1}y'$ so that $Q(x) = Q(y'') = 0$; $B(x, y'') = 1$. With no loss of generality, we may assume that $y'' = y$; thus
$$Q(ax + by) = a^2Q(x) + b^2Q(y) + abB(x, y) = ab$$
for all $a, b \in F$. Clearly $\langle x, y\rangle \cap \langle x, y\rangle^\perp = 0$ and so
$$V = \langle x, y\rangle \oplus \langle x, y\rangle^\perp.$$
If $z, \in \langle x, y\rangle^\perp$ then
$$Q(ax + by + z) = Q(ax + by) + Q(z) + B(ax + by, z) = ab + Q(z)$$
so (a) holds. Conclusion (b) is clear. □

Every quadratic extension $E \supset F$ of finite fields[1] (so that $[E : F] = 2$) naturally gives rise to a quadratic form $E \to F$, as we now describe. There exists a nonidentity field automorphism of E, which we denote $z \mapsto \overline{z}$, fixing every element of F (in other words the extension $E \supset F$ is Galois). The norm map
$$N : E \to F, \quad z \mapsto z\overline{z}$$

[1]This much is true more generally for any separable extension of degree 2. It is the field-theoretic analogue of the fact from group theory that any subgroup of index 2 is normal.

is a quadratic form with no absolute points (since if $z \neq 0$ then $N(z) = z\overline{z} \neq 0$). If we choose a basis $\{\alpha, \beta\}$ for E over F, then

$$N(r\alpha + s\beta) = (r\alpha + s\beta)(r\overline{\alpha} + s\overline{\beta}) = ar^2 + brs + cs^2$$

where $aX^2 + bX + c \in F[X]$ is irreducible of degree 2. This fact has a useful converse, as we now observe:

23.6 Lemma. A quadratic polynomial $f(X) = aX^2 + bX + c \in \mathbb{F}_q[X]$ is irreducible over $\mathbb{F}_q$ iff there exists a basis $\{\alpha, \beta\}$ for $\mathbb{F}_{q^2}$ over $\mathbb{F}_q$, such that $N(r\alpha + s\beta) = ar^2 + brs + cs^2$.

Proof. There exists $\omega \in \mathbb{F}_{q^2}$ such that $f(\omega) = 0$. Since the norm map is surjective (see Appendix A1), there exists $\gamma \in \mathbb{F}_{q^2}$ such that $N(\gamma) = a$. Now

$$N(r\gamma - s\gamma\omega) = aN(r - s\omega) = a(r - s\omega)(r - s\overline{\omega}) = ar^2 + brs + cs^2$$

as required. The converse follows by the preceding remarks. □

We now consider the isometry group $O(V, Q)$ of a quadratic form $Q : V \to F$, which is typically quite large. The primary source for this assertion is the following result.

23.7 Theorem (Witt). Let $Q : V \to F$ be a quadratic form, and let $\theta : U \to \theta(U)$ be an isometry between two subspaces of V, i.e. $Q(\theta(u)) = Q(u)$ for all $u \in U$. Then the following two conditions are equivalent.

(i) $\theta(U \cap V^\perp) = \theta(U) \cap V^\perp$.

(ii) θ extends to an isometry of V to itself, i.e. there is an isometry $\widehat{\theta} : V \to V$ such that the restriction $\widehat{\theta}\big|_U = \theta$.

The proof of this result, given below, is rather technical; we recommend that on first reading, the student skip it. However, the reader is asked to consider some of the important implications of this result, including the following.

23.8 Corollary. Let $Q : V \to F$ be a nondegenerate quadratic form.

(a) Any two maximal totally singular subspaces have the same dimension.

(b) Any two singular points lie in the same number of maximal totally singular subspaces.

Proof. Let $U, U' \leq V$ be maximal totally singular subspaces, and assume $\dim U \leq \dim U'$. Let $\theta : U \to U'$ be any injective linear transformation. Since Q vanishes on both U and U', θ is an isometry. By Witt's Theorem 23.7 there exists an isometry $\widehat{\theta} : V \to V$ extending θ. Since U is maximal, so is $\widehat{\theta}(U) = \theta(U) \leq U'$, and this forces equality: $\theta(U) = U'$. In particular $\dim U = \dim \theta(U) = \dim U'$ so (a) holds.

If $\langle x \rangle$ and $\langle y \rangle$ are singular points then we have an isometry $\theta : \langle x \rangle \to \langle y \rangle$ defined by $\theta(cx) = cy$ for all $c \in F$. By Witt's Theorem this extends to an isometry $\widehat{\theta} : V \to V$ which induces a bijection between the maximal totally singular subspaces containing x, and those containing y. This gives (b). □

Proof of Theorem 23.7. Assuming (ii) then (i) must follow since the isometry $\widehat{\theta} : V \to V$ must preserve the radical $V^\perp$. Thus we assume (i) and seek to prove (ii). We first show that without loss of generality,

$$U \supset V^\perp \text{ and } \theta(U) \supset V^\perp. \tag{23.9}$$

To show this, let $W \subseteq V^\perp$ such that $V^\perp = W \oplus (U \cap V^\perp) = W \oplus (\theta(U) \cap V^\perp)$; such a subspace exists by Exercise 4. It follows that

$$U + V^\perp = U \oplus W \qquad \text{and} \qquad \theta(U) + V^\perp = \theta(U) \oplus W.$$

We extend θ to the map

$$U \oplus W \to \theta(U) \oplus W, \qquad u + w \mapsto \theta(u) + w$$

for all $u \in U$, $w \in W$. This is an isometry since

$$\begin{aligned} Q(\theta(u) + w) &= Q(\theta(u)) + Q(w) \qquad \text{since } B(\theta(u), w) = 0 \\ &= Q(u) + Q(w) \\ &= Q(u + w) \qquad \text{since } B(u, w) = 0 \end{aligned}$$

so we may replace U by $U + V^\perp$, and $\theta(U)$ by $\theta(U) + V^\perp$ if necessary, so that both U and $\theta(U)$ contain the radical $V^\perp$. Furthermore we may assume the containment $U \supset V^\perp$ is proper; otherwise the argument above applies with $W = 0$. So we may assume that (23.9) holds.

We proceed to verify the conclusion by induction on $k = \dim(U/V^\perp)$. The initial case $k = 0$ was settled above; so now we take $k \geq 1$. Let $H < U$ be a subspace of codimension 1 containing $V^\perp$, so that the subspace $\theta(H) < \theta(U)$ of codimension 1 is isometric to H and also contains $V^\perp$. We may assume that

$$\theta \text{ fixes every element of } H. \tag{23.10}$$

Otherwise by the inductive hypothesis, the restriction $\theta\big|_H : H \to \theta(H)$ extends to an isometry $\widehat{\theta} : V \to V$. The isometry $\psi = \widehat{\theta}^{-1} \circ \theta : U \to \psi(U)$ does fix every element of H. If we can extend ψ to an isometry $\widehat{\psi} : V \to V$, then the isometry $\widehat{\theta} \circ \widehat{\psi} : V \to V$ extends $\theta : U \mapsto \theta(U)$. So it suffices to assume (23.10) holds.

Now both U and $\theta(U)$ contain H as a subspace of codimension 1 fixed elementwise by θ. There exists $z \in U$ such that

$$U = H \oplus \langle z \rangle; \quad \theta(U) = H \oplus \langle \theta(z) \rangle.$$

We assume $\theta(z) \neq z$; otherwise the conclusion follows. For all $v \in H$ we have

$$B(z - \theta(z), v) = B(z, v) - B(\theta(z), v) = B(\theta(z), \theta(v)) - B(\theta(z), v) = 0$$

since $\theta(v) = v$; thus $H \subseteq y^\perp$ where $y = z - \theta(z)$.

(23.11) For all $x \in U$ we have $\theta(x) - x \in \langle y \rangle$;

this follows since $x \in U = H \oplus \langle y \rangle$ where θ fixes every element of H. We may now assume that

(23.12) $U, \theta(U) \subseteq y^\perp$. In particular, $y \in y^\perp$.

For suppose $U \not\subseteq y^\perp$. Then there exists $x \in U$ such that

$$\begin{aligned} 0 \neq B(x, y) &= B(x, z) - B(x, \theta(z)) \\ &= B(\theta(x), \theta(z)) - B(x, \theta(z)) = B(\theta(x) - x, \theta(z)) \end{aligned}$$

where $\theta(x) - x \in \langle y \rangle$, so that $\theta(U) \not\subseteq y^\perp$. Now there exists a subspace W such that $y^\perp = H \oplus W$. It follows that $V = U \oplus W$. (Since $U \cap W \subseteq U \cap y^\perp = H$ we have $U \cap W \subseteq H \cap W = 0$.) The mapping $V \to V$ defined by $u + w \mapsto \theta(u) + w$ (for $u \in U$, $w \in W$) maps U to $\theta(U)$ and is an isometry since

$$\begin{aligned} Q(\theta(u) + w) &= Q(\theta(u)) + Q(w) + B(\theta(u), w) \\ &= Q(u) + Q(w) + B(u, w) \\ &= Q(u + w) \end{aligned}$$

since $W \subseteq y^\perp$. Thus we may assume (23.12). Next we may assume

(23.13) $\theta(U) = U = y^\perp < V$.

For if $\theta(U) = U$ then we may write $y^\perp = U \oplus S$ and an extension of θ to $y^\perp$ is given by

$$u + s \mapsto \theta(u) + s$$

for all $u \in U$, $s \in S$; this is an isometry since

$$\begin{aligned} Q(\theta(u)+s) &= Q(\theta(u)) + Q(s) + B(\theta(u), s) \\ &= Q(u) + Q(s) + B(u, s) \qquad \text{since } u - \theta(u) \in \langle y \rangle \subseteq S^\perp \\ &= Q(u+s). \end{aligned}$$

Replacing U by $y^\perp$ gives (23.13) in this case.

On the other hand if $\theta(U) \neq U$ then we choose vectors d, e such that $U = H \oplus \langle d \rangle$ and $\theta(U) = H \oplus \langle e \rangle$. There exists a subspace W such that $y^\perp = (U + \theta(U)) \oplus W$. Consider the subspace $S = W \oplus \langle d+e \rangle$. Since $U + \theta(U) = U \oplus \langle d+e \rangle = \theta(U) \oplus \langle d+e \rangle$, we have

$$y^\perp = U \oplus S = \theta(U) \oplus S.$$

As before, an isometry extending θ to $y^\perp$ is given by $u + s \mapsto \theta(u) + s$ for all $u \in U$, $s \in S$. Once again we may replace both U and $\theta(U)$ by $y^\perp$; so in any case we may assume that (23.13) holds.

Since $B(z, \theta(z)) = B(z, z-y) = B(z, z)$, we have

$$Q(y) = Q(z-\theta(z)) = Q(z) + Q(\theta(z)) - B(z, \theta(z)) = 2Q(z) - B(z, z) = 0.$$

There exists $y' \in H^\perp$ such that $\{y, y'\}$ is a hyperbolic pair, by Lemma 23.5 applied to $Q\big|_{H^\perp}$; similarly there exists $y'' \in H^\perp$ such that $\{\theta(y), y''\}$ is a hyperbolic pair. Now

$$V = H \oplus \langle y, y' \rangle = H \oplus \langle \theta(y), y'' \rangle$$

where $\langle y, y' \rangle^\perp = \langle \theta(y), y'' \rangle^\perp = H$. An extension of θ to V is given by

$$ay + by' + u \mapsto a\theta(y) + by'' + u$$

for all $a, b \in F$, $u \in H$; this is an isometry since

$$Q(ay + by' + u) = ab + Q(u) = Q(a\theta(y) + by'' + u). \qquad \square$$

Exercises 23.

1. Classify all isometry classes of quadratic forms on $\mathbb{C}^n$, as we have done for $\mathbb{R}^n$. Justify your answer.
2. Let $Q : V \to F$ be a quadratic form. Show that if $\theta \in O(V, Q)$ is an isometry and $\theta(U) = U$ for some subspace $U \leq V$, then $\theta(U^\perp) = U^\perp$.
3. Let $Q : V \to F$ be a quadratic form with associated bilinear form B, and suppose $u \in V$ such that $Q(v) \neq 0$. Define $R_u : V \to V$ by $v \mapsto v - \frac{B(u,v)}{Q(v)} v$. Show that R_u is an isometry fixing every element of $u^\perp$ and mapping $u \mapsto -u$.
4. Let V be an n-dimensional vector space over F and let $U, U' \leq V$ be k-dimensional subspaces. Show that there exists a subspace $W \leq V$ such that $V = U \oplus W = U' \oplus W$.

Hint: Let $v_1, \ldots, v_r$ be a basis for $U \cap U'$. Extend to a basis $v_1, \ldots, v_r, v_{r+1}, \ldots, v_k$ for U and a basis $v_1, \ldots, v_r, v'_{r+1}, \ldots, v'_k$ for U'. Show that the union of the preceding bases gives a basis for $U + U'$, and adjoin new vectors $v_{k+1}, \ldots, v_n$ to obtain a basis for V. Consider the subspace spanned by $v_{r+1}+v'_{r+1}, \ldots, v_k+v'_k, v_{k+1}, \ldots, v_n$.

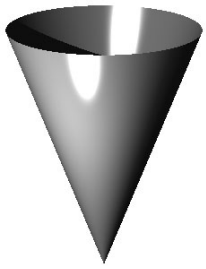
5. Let $Q : V \to F$ be a quadratic form on a 2-dimensional vector space V. Show that either 0, 1, 2 or all of the points of V are singular. Give examples to show that each of these four possibilities may actually occur. Can all these possibilities occur for every choice of field F? Explain.

24. Quadrics and Polar Spaces

Let V be an n-dimensional vector space over a field F. A **quadric** in $\mathbb{P}V = \mathbb{P}^{n-1}(F)$ may be defined as the set of singular points of a quadratic form $Q : V \to F$. Such a quadric is **nondegenerate** if Q is. Note that a quadric in a projective plane is simply a conic. Three familiar quadrics in real projective 3-space are the *elliptic* and *hyperbolic quadrics,* and the *quadratic cone,* as shown below. Note that maximal subspaces lying on the quadric (i.e. maximal totally singular subspaces) are points in the case of the elliptic quadric; or lines in the case of the hyperbolic quadric or the cone.

Elliptic Quadric
$x_1^2 + x_2^2 + x_3^2 - x_4^2 = 0$

Hyperbolic Quadric
$x_1x_2 + x_3x_4 = 0$



Cone
$x_2^2 + x_3x_4 = 0$

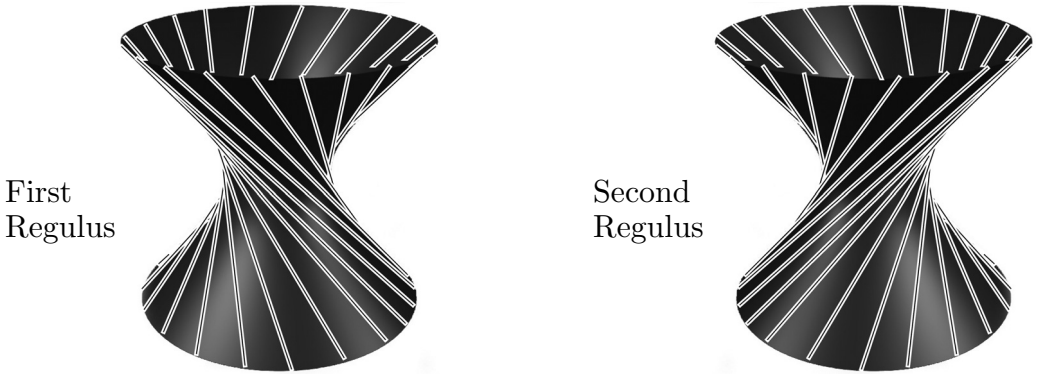
These three quadrics are distinguished by the fact that the elliptic quadric contains no lines; the cone contains infinitely many lines, all passing through a common point (the vertex of the cone); and the hyperbolic quadric contains two infinite families of mutually disjoint lines. (see below). In fact the hyperbolic quadric contains a family of lines $\{\ell_s : s \in \mathbb{R} \cup \{\infty\}\}$ (called a **regulus**) defined by

$$\ell_s = \langle (1, 0, -s, 0),\ (0, s, 0, 1) \rangle \text{ for } s \in \mathbb{R}; \quad \text{and } \ell_\infty = \langle (0, 1, 0, 0),\ (0, 0, 1, 0) \rangle$$

which partitions its points; and a second regulus $\{\ell'_s : s \in \mathbb{R} \cup \{\infty\}\}$, described as the **alternate regulus** to the first regulus, given by

$$\ell'_s = \langle (1, 0, 0, -s),\ (0, s, 1, 0) \rangle \text{ for } s \in \mathbb{R}; \quad \text{and } \ell'_\infty = \langle (0, 1, 0, 0),\ (0, 0, 0, 1) \rangle$$

The two **reguli** (plural of *regulus*) are as shown:



Topologically, the elliptic and hyperbolic quadrics in $\mathbb{P}^3(\mathbb{R})$ are homeomorphic to the sphere S^2 and the torus T^2 respectively.

If we disregard the equation $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 0$ (which has no real solutions, yielding the empty quadric) then every nondegenerate quadric in $\mathbb{P}^3(\mathbb{R})$ is equivalent to one with equation either

$$x_1^2 + x_2^2 + x_3^2 - x_4^2 = 0 \qquad \text{or} \qquad x_1^2 + x_2^2 - x_2^3 - x_4^2 = 0.$$

The latter equation defines a hyperbolic quadric since an invertible linear change of variables

$$(x_1,\, x_2,\, x_3,\, x_4) = (y_1{+}y_2,\, y_3{+}y_4,\, y_1{-}y_2,\, y_3{-}y_4)$$

transforms the equation $x_1^2 + x_2^2 - x_3^2 - x_4^2 = 0$ into the equation $y_1y_2 + y_3y_4 = 0$.

Just as a single equivalence class of nondegenerate conics in the real projective plane gives rise to several inequivalent real affine (i.e. Euclidean) plane conics, namely ellipses, parabolas and hyperbolas, so also our two nondegenerate quadrics in $\mathbb{P}^3(\mathbb{R})$ give rise to many different quadrics in Euclidean 3-space $\mathbb{A}^3(\mathbb{R})$. For example the affine *hyperbolic paraboloid* $z = x^2 - y^2$, when homogenized by the addition of a fourth variable w, becomes $x^2 - y^2 - zw = 0$; then setting

$$(x, y, z, w) = (x_1, x_3, x_4{-}x_2, x_2{+}x_4)$$

transforms the equation to $x_1^2{+}x_2^2{-}x_3^2{-}x_4^2 = 0$. If we then set $(x_1, x_2, x_3, x_4) = (X, Y, Z, 1)$ then we obtain the *one-sheeted hyperboloid* $Z^2 = X^2 + Y^2 - 1$.

We see from the classification above that $\mathbb{P}^{n-1}(\mathbb{R})$ has $\lfloor \frac{n}{2} \rfloor$ equivalence classes of (nonempty) nondegenerate quadrics; note that this number tends to ∞ as $n \to \infty$. By contrast, in $\mathbb{P}^{n-1}(\mathbb{F}_q)$ there are just two types of nondegenerate quadrics for n even (the *elliptic* and *hyperbolic quadric*) and one type of nondegenerate quadric for n odd (the *parabolic type*). The main property of $\mathbb{R}$ which accounts for this difference is the fact that $\mathbb{R}$ is an ordered field, whereas finite fields are not ordered.

We proceed to classify quadrics over finite fields. Here some aspects of the discussion are more delicate for fields of even characteristic, and the beginning student may prefer to ignore this case on first reading. But fields of even characteristic are of such significance and interest that the extra effort to understand them proves rewarding. Regardless of the characteristic we require a more complete description of the quadratic form and its relationship to an associated bilinear form.

Note that singular points are simply points on the associated quadric; more generally, totally singular subspaces are simply projective subspaces contained in the quadric.

This leads to the desired classification of quadratic forms on $\mathbb{F}_q^n$:

24.1 Theorem. Let V be an $(n{+}1)$-dimensional vector space over $\mathbb{F}_q$, and let $Q : V \to \mathbb{F}_q$ be a nondegenerate quadratic form. Fix $a, b, c \in \mathbb{F}_q$ such that the quadratic polynomial $aX^2 + bX + c \in \mathbb{F}_q[X]$ is irreducible over $\mathbb{F}_q$. If q is odd, fix a nonsquare $\eta \in \mathbb{F}_q$.

If $n{=}2k$ then after a suitable linear change of coordinates, Q is given by either

(H) $x_1x_2 + x_3x_4 + \cdots + x_{2k-1}x_{2k}$; or

(E) $x_1x_2 + x_3x_4 + \cdots + x_{2k-3}x_{2k-2} + ax_{2k-1}^2 + bx_{2k-1}x_{2k} + cx_{2k}^2$.

If $n = 2k{+}1$ then after a suitable linear change of coordinates, Q is given by either

(P) $x_1x_2 + x_3x_4 + \cdots + x_{2k-1}x_{2k} + x_{2k+1}^2$; or

(P′) $x_1x_2 + x_3x_4 + \cdots + x_{2k-1}x_{2k} + \eta x_{2k+1}^2$. (This case arises only for q odd).

Quadratic forms as in (H) and (E), and the associated quadrics, are of **hyperbolic** and **elliptic** type respectively. Quadrics as in (P) or (P′), and the associated quadrics, are of **parabolic**) type. Denoting the quadratic forms in (P) and (P′) by Q and Q' respectively, we observe that Q is isometric to $\eta Q'$ since

$$Q(x_1, \eta x_2, x_3, \eta x_4, \cdots, x_{2k-1}, \eta x_{2k}, x_{2k+1}) = \eta Q(x_1, x_2, x_3, x_4, \cdots, x_{2k-1}, x_{2k}, x_{2k+1})$$

and so the forms Q and Q' have the same singular points; accordingly their corresponding quadrics have the same geometric properties.

Proof of Theorem 24.1. If $n = 1$ then $Q(x_1) = cx_1^2$ where $c \in \mathbb{F}_q$ is nonzero. If $c = t^2$ for some nonzero $t \in \mathbb{F}_q$ then the change of variable $y_1 = tx_1$ yields $cx_1^2 = y_1^2$ and (P) holds. Otherwise q is odd and $c \in \mathbb{F}_q$ is a nonsquare so that $c = \eta t^2$ for some nonzero $t \in \mathbb{F}_q$; but then the change of variable $y_1 = tx_1$ yields $cx_1^2 = \eta y_1^2$ so (P′) holds.

Suppose $n = 2$. If there is a singular point $\langle x \rangle$ then by Lemma 23.5 the quadratic form may be written as $Q(x_1, x_2) = x_1x_2$ after a suitable coordinate change, so that (H) holds. So we may assume there is no singular point. By Lemma 23.7 there exists a basis $\{\alpha, \beta\}$ for $\mathbb{F}_{q^2}$ over $\mathbb{F}_q$ such that the norm map $N : \mathbb{F}_{q^2} \to \mathbb{F}_q$, $z \mapsto z^{q+1}$ satisfies $N(r\alpha + s\beta) = ar^2 + brs + cs^2$. We may similarly identify V with $\mathbb{F}_{q^2}$, and choose a basis

$\{\alpha', \beta'\}$ for $\mathbb{F}_{q^2}$ over $\mathbb{F}_q$, in such a way that $Q : V \to F$ satisfies $Q(x_1, x_2) = N(x_1\alpha' + x_2\beta')$. A change of basis from $\{\alpha', \beta'\}$ to $\{\alpha, \beta\}$ shows that $Q(x_1, x_2)$ is equivalent to $ar^2+brs+cs^2$, and so (E) holds.

Finally suppose $n \geq 3$. By the Chevalley-Warning Theorem A2.3 (Appendix 2), the number of solutions $(x_1, x_2, \ldots, x_n) \in \mathbb{F}_q^n$ to the equation $Q(x) = 0$ is divisible by p where $q = p^e$; since the zero vector is a solution, there must exist a nonzero solution. Thus there exists a singular point $\langle u \rangle$. By Lemma 23.5 there exists another singular point $\langle v \rangle$ such that the restriction of Q to the $(n-2)$-space $\langle u, v \rangle^\perp$ is nondegenerate, and

$$Q(x_1u + x_2v + z) = x_1x_2 + Q(z)$$

for all $x_1, x_2 \in \mathbb{F}_q$ and $z \in \langle u, v \rangle^\perp$. The result follows by induction. □

24.2 Corollary. Let $Q : V \to \mathbb{F}_q$ be a nondegenerate quadratic form on a finite-dimensional vector space over a field $\mathbb{F}_q$ where q is odd. Then the isometry type of Q is uniquely determined by the quantities q, $n = \dim V$, and $\operatorname{disc} Q$. Assuming $n = 2k$, the form Q is hyperbolic or elliptic according as $(-1)^k \operatorname{disc} Q$ is a square or a nonsquare.

Proof. Suppose first that $n = 2k$. By Theorem 24.1, to within an isometry the matrix of B is given by either

(H) $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \oplus \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \oplus \cdots \oplus \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ (k summands), or

(E) $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \oplus \cdots \oplus \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \oplus \left[\begin{smallmatrix} a & b/2 \\ b/2 & c \end{smallmatrix}\right]$ (k summands, where aX^2+bX+c is irreducible over $\mathbb{F}_q$).

The determinant of this matrix is $(-1)^k$ in case (H), or $(-1)^k(b^2 - 4ac)/4$ in case (E). Since $aX^2+bXY + cY^2$ is irreducible, its discriminant $b^2 - 4ac$ is a nonsquare in $\mathbb{F}_q$ and the result follows.

Now suppose $n = 2k+1$. Again by Theorem 24.1, to within an isometry the matrix of B is given by either

(P) $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \oplus \cdots \oplus \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \oplus [1]$ ($k+1$ summands), or

(P′) $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \oplus \cdots \oplus \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \oplus [\eta]$ ($k+1$ summands)

where $\eta \in \mathbb{F}_q$ is a nonsquare. The determinant of this matrix is $(-1)^k$ in case (P), or $(-1)^k\eta$ in case (P′). One of these values is a square and the other is a nonsquare, so once again the result follows. □

24.3 Theorem. Let V be an n-dimensional vector space over $\mathbb{F}_q$, and let $Q : V \to F$ be a nondegenerate quadratic form. Suppose $\langle x\rangle$ is a singular point. Then Q naturally induces a nondegenerate quadratic form on the $(n-2)$-space $x^\perp/\langle x\rangle$. This form has the same type (hyperbolic, parabolic or elliptic) as the original space (unless $n = 2$ in which case $x^\perp/\langle x\rangle = 0$).

Proof. Since $\langle x\rangle$ is the radical of $x^\perp$, by Lemma 23.4 the map

$$\widetilde{Q} : x^\perp/\langle x\rangle \to F, \qquad v + \langle x\rangle \mapsto Q(v)$$

is a well-defined quadratic form induced naturally by Q. There is no loss of generality in assuming Q is given by one of the four canonical forms listed in Theorem 24.1; in particular

$$Q(x) = x_1x_2 + Q_1(x_3, x_4, \ldots, x_n)$$

using coordinates $(x_1, x_2, \ldots, x_n)$ for V relative to some basis $e_1, e_2, \ldots, e_n$, where Q_1 has the same type as Q (H, E, P or P′) but in 2 fewer variables. By Witt's Theorem 23.7 there is no loss of generality in assuming that $x = e_1$. Then $x^\perp$ has basis $\{e_1, e_3, e_4, e_5, \ldots, e_n\}$ and $x^\perp/\langle x\rangle$ has basis $\{e_i+\langle e_1\rangle \,:\, i = 3, 4, 5, \ldots, n\}$. The quadratic form induced on $x^\perp/\langle x\rangle$ is therefore Q_1. □

24.4 Theorem. Let V_1 and V_2 be even-dimensional vector spaces over $\mathbb{F}_q$. Then the orthogonal direct sum $V_1 \perp V_2$ is

(i) hyperbolic, if V_1 and V_2 are either both hyperbolic or both elliptic; or

(ii) elliptic, if one of V_1, V_2 is hyperbolic and the other elliptic.

Proof. Let $\dim V_i = 2k_i$. For q odd we have $\operatorname{disc} Q = (\operatorname{disc} Q_1)(\operatorname{disc} Q_2)$ and so

$$(-1)^{k_1+k_2} \operatorname{disc} Q = (-1)^{k_1} \operatorname{disc} Q_1 \cdot (-1)^{k_2} \operatorname{disc} Q_2$$

where $(-1)^{k_i} \operatorname{disc} Q_i = 1$ or -1 according as Q_i is hyperbolic or elliptic, and similarly for Q. The result follows in this case.

Hence we may assume q is even. It clearly suffices to assume $k_1 = k_2 = 1$ with both Q_1 and Q_2 elliptic; the general case then follows by taking orthogonal direct sums with sufficiently many hyperbolic lines. We may assume

$$Q_1(x_1, x_2) = ax_1^2 + bx_1x_2 + cx_2^2, \qquad Q_2(y_1, y_2) = ay_1^2 + by_1y_2 + cy_2^2$$

where at^2+bt+c is irreducible over $\mathbb{F}_q$; note that $abc \neq 0$ in this case. Since

$$(24.5)\qquad \begin{aligned} Q(x_1,x_2,y_1,y_2) &= ax_1^2{+}bx_1x_2{+}cx_2^2 \;+\; ay_1^2{+}by_1y_2{+}cy_2^2 \\ &= (x_1{+}y_1)(ax_1{+}ay_1{+}bx_2) + (x_2{+}y_2)(cx_2{+}cy_2{+}by_1) \end{aligned}$$

which has the form $u_1u_2 + u_3u_4$ where the u_i's are the latter parenthesized quantities in (24.5) and the linear change of variables $(x_1,x_2,y_1,y_2) \mapsto (u_1,u_2,u_3,u_4)$ is nonsingular. So (24.5) is a quadratic form of hyperbolic type as required. □

Let $Q : V \to \mathbb{F}_q$ be a nondegenerate quadratic form where V is n-dimensional over $\mathbb{F}_q$. It is customary to denote the isometry type of (V,Q) as

$O^+_{2k}(q)$ if $n{=}2k$ and Q is hyperbolic;

$O^-_{2k}(q)$ if $n{=}2k$ and Q is elliptic; or

$O_{2k+1}(q)$ if $n{=}2k{+}1$, in which case Q is parabolic.

We summarize the statement of Theorem 24.4 as

$$O^{\varepsilon}_{2k}(q) \;\perp\; O^{\varepsilon'}_{2k'}(q) \;\simeq\; O^{\varepsilon\varepsilon'}_{2(k+k')}(q)\,.$$

24.6 Theorem. The number of singular points and maximal totally singular subspaces for each type of nondegenerate quadratic form over $\mathbb{F}_q$ are as listed in the table to the right.

isometry type	no. of singular points	no. of maximal totally singular subspaces
$O^+_{2k}(q)$	$\dfrac{q^k-1}{q-1}\big(q^{k-1}+1\big)$	$2(q{+}1)(q^2{+}1)\cdots(q^{k-1}{+}1)$
$O^-_{2k}(q)$	$\dfrac{q^{k-1}-1}{q-1}\big(q^k+1\big)$	$(q^2{+}1)(q^3{+}1)\cdots(q^k{+}1)$
$O_{2k+1}(q)$	$\dfrac{q^{2k}-1}{q-1}$	$(q{+}1)(q^2{+}1)\cdots(q^k{+}1)$

Proof. For each $a \in \mathbb{F}_q$, let $N_k(a)$ be the set of vectors $(x_1,x_2,\dots,x_{2k}) \in \mathbb{F}_q^{2k}$ such that

$$x_1x_2 + x_3x_4 + \cdots + x_{2k-1}x_{2k} = a.$$

It is easy to see that $|N_k(a)| = |N_k(1)|$ for all nonzero $a \in \mathbb{F}_q$ using the bijection

$$N_k(1) \to N_k(a), \quad (x_1,x_2,x_3,x_4,\dots,x_{2k-1},x_{2k}) \mapsto (x_1,ax_2,x_3,ax_4,\dots,x_{2k-1},ax_{2k}).$$

Writing $n_k = |N_k(0)|$, it follows that $|N_k(a)| = \frac{q^{2k}-n_k}{q-1}$ for $a \neq 0$. We claim that in fact

$$(24.7)\qquad |N_k(0)| = n_k = q^{2k-1}{+}q^k{-}q^{k-1}\text{; and consequently } |N_k(a)| = q^{k-1}(q^k{-}1) \text{ for all nonzero } a \in \mathbb{F}_q\,.$$

Vectors in $N_k(0)$ are partitioned into those solutions with $x_1 = 0$ and $x_2 \in \mathbb{F}_q$ arbitrary, with n_{k-1} choices of the remaining coordinates; and those solutions with $x_1 \neq 0$, the coordinates $x_3, x_4, \ldots, x_{2k}$ arbitrary and $x_2 \in \mathbb{F}_q$ uniquely determined. This gives

$$n_k = qn_{k-1} + (q-1)q^{2k-2}.$$

Taking either $n_0 = 1$ or $n_1 = 2q-1$ as the initial case, (24.7) follows easily by induction.

We may assume the nondegenerate quadratic form Q is the standard representative of its isometry class as given in Theorem 24.1. In the case of $O^+_{2k}(q)$ the form $Q(x) = x_1x_2 + x_3x_4 + \cdots + x_{2k-1}x_{2k}$ has

$$\frac{n_k - 1}{q-1} = \frac{q^k - 1}{q-1}(q^{k-1} + 1)$$

singular points, arising from the $n_k - 1$ nonzero vectors in $N_k(0)$. For $O^-_{2k}(q)$ the form $Q(x) = x_1x_2 + \cdots + x_{2k-3}x_{2k-2} + ax^2_{2k-1} + bx_{2k-1}x_{2k} + cx^2_{2k}$ (with $at^2 + bt + c$ irreducible over $\mathbb{F}_q$) has

$$|N_{k-1}(0)| + (q^2-1)|N_{k-1}(1)| = q^{2k-1} - q^k + q^{k-1}$$

singular vectors, counting separately those with $(x_{2k-1}, x_{2k}) = (0,0)$ or $\neq (0,0)$, giving

$$\frac{q^{2k-1} - q^k + q^{k-1} - 1}{q-1} = \frac{q^{k-1} - 1}{q-1}(q^k + 1)$$

as the required number of singular points. Similarly in the $O_{2k+1}(q)$ case $Q(x) = x_1x_2 + \cdots + x_{2k-1}x_{2k} + x^2_{2k+1}$ has

$$|N_k(0)| + (q-1)|N_k(1)| = q^{2k}$$

singular vectors and hence $\frac{q^{2k}-1}{q-1}$ singular points.

In each of the three cases we denote by m_k the number of maximal totally singular subspaces. This number is found by first counting in two different ways the number of pairs $(\langle x\rangle, U)$ where U is a totally singular k-space containing a point $\langle x\rangle$. Consider first the case $O^+_{2k}(q)$ in which every such U is k-dimensional. There are m_k choices of U, each containing $\begin{bmatrix} k \\ 1 \end{bmatrix}_q = (q^k - 1)/(q-1)$ points. On the other hand there are

$$\frac{q^k - 1}{q-1}\big(q^{k-1} + 1\big)$$

choices of singular point $\langle x\rangle$; and for each, the totally singular k-subspaces U containing x are in one-to-one correspondence with the m_{k-1} totally singular $(k-1)$-subspaces of the $O^+_{2(k-1)}(q)$-space $x^\perp/\langle x\rangle$. Thus

$$\frac{q^k - 1}{q-1}m_k = \frac{q^k - 1}{q-1}\big(q^{k-1} + 1\big)m_{k-1}$$

which yields the recurrence relation $m_k = (q^{k-1}+1)m_{k-1}$. Given that an $O_2^+(q)$-space has 2 singular points as its maximal totally singular subspaces, we have $m_1 = 2$ and so by induction

$$m_k = 2(q+1)(q^2+1)\cdots(q^{k-1}+1)$$

as claimed.

In the $O_{2k}^-(q)$ case, maximal totally singular subspaces have dimension $k-1$ and there are $\frac{q^{k-1}-1}{q-1}\big(q^k+1\big)$ choices of the singular point $\langle x\rangle$, so the same reasoning gives

$$\frac{q^{k-1}-1}{q-1}m_k = \frac{q^{k-1}-1}{q-1}\big(q^k+1\big)m_{k-1}$$

whence $m_k = (q^k+1)m_{k-1}$ in this case. Since $O_2^-(q)$ has $\{0\}$ as its unique maximal totally singular subspace, we have $m_1 = 1$ and so by induction

$$m_k = \prod_{2\le i\le k}(q^i+1) = (q^2+1)(q^3+1)\cdots(q^k+1).$$

Finally in the $O_{2k+1}(q)$ case, maximal totally singular subspaces have dimension k and there are $(q^{2k}-1)/(q-1)$ choices of the singular point $\langle x\rangle$, so

$$\frac{q^k-1}{q-1}m_k = \frac{q^{2k}-1}{q-1}m_{k-1}$$

which gives $m_k = (q^k+1)m_{k-1}$. Since the $O_3(q)$-quadric is a conic with $q+1$ points as its maximal totally singular subspaces, we have $m_1 = q+1$ so by induction

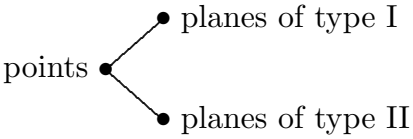
$$m_k = (q+1)(q^2+1)\cdots(q^k+1).$$

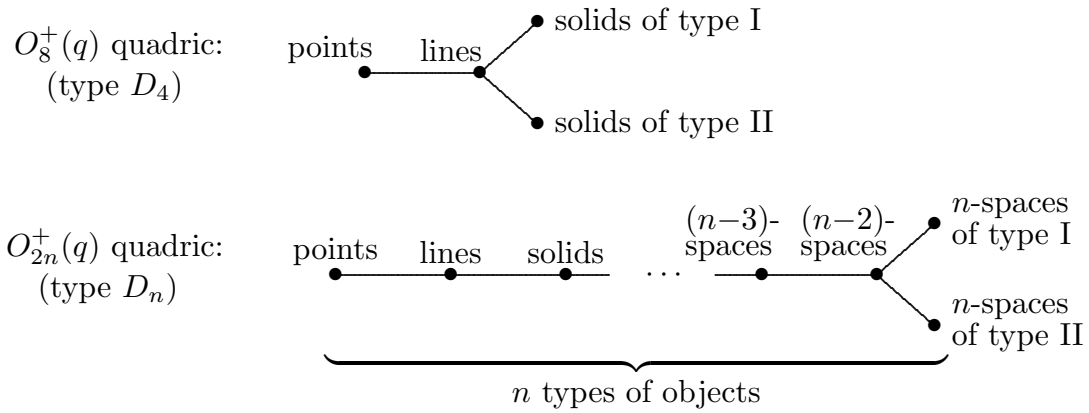
□

The incidence system formed by all totally singular subspaces with respect to a hyperbolic quadratic form over $\mathbb{F}_q$, i.e. the subspaces contained in an $O_{2k}^+(q)$ quadric, are described by Coxeter-Dynkin diagrams of type D_k:

$O_4^+(q)$ quadric: (type D_2)
- lines of type I
- lines of type II

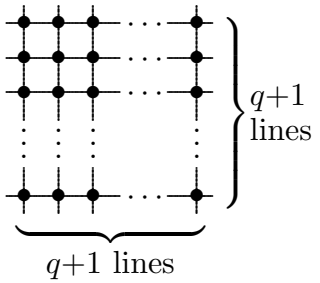
$O_6^+(q)$ quadric: (type D_3)





As with the diagrams of type A_n for projective n-spaces, the D_n diagrams are endowed with many layers of meaning. For example, the evident symmetry of the diagram represents the existence of an automorphism interchanging the two types of totally singular n-spaces. More insight gleaned from the D_n diagrams will emerge as we proceed to examine some particular cases.

24.8 The $O_4^+(q)$-Quadric. The $O_4^+(q)$ quadric consists of $(q+1)^2$ points and $2(q+1)$ lines, forming a grid as shown. There are two types of lines, each type forming a *regulus.* Any two lines in the same regulus are skew; but every line in one regulus meets every line in the opposite ('alternate') regulus. We say that two lines are *incident* if they meet in a point, i.e. if they are from opposite reguli. (This is a different definition of incidence than the relation of inclusion used in projective space!) Since every line of one regulus is incident with every line of the alternate regulus, the geometry is simply the generalized digon that we have encountered previously. It is no coincidence that the D_2 diagram coincides with the $A_1 \oplus A_1$ diagram previously used for this geometry.



The student may at first wonder why the diagram contains no node representing points of the quadric. In fact such a node is not needed, for reasons that we hope will become clear in time. For now the reader may observe that the geometry (in this case, generalized digon) formed by the two families of lines of the quadric has as its incidence graph a complete bipartite graph $K_{q+1,q+1}$, and that points of the quadric are adequately represented as the $(q+1)^2$ edges of this graph.

24.9 The $O_{2n}^+(q)$-Quadric. In just the same way, the O_{2n}^+-quadric contains two types of totally singular n-spaces, in fact $(q+1)(q^2+1)\cdots(q^{n-1}+1)$ of each type; this accounts for the factor of 2 appearing in the formula from Theorem 24.6. The geometry of the quadric contains n types of objects: the totally singular subspaces of dimension 1, 2, ..., $n-2$, and the two types of totally singular n-subspaces. It does not matter which type

of totally singular n-subspace is designated as type I (and the other as type II) since an automorphism of the geometry interchanges these two types. Every totally singular $(n-1)$-space U lies in exactly one totally singular n-subspace of each type, corresponding to the two singular points of $U^\perp/U \simeq O_2^+(q)$. Accordingly we say that two totally singular n-subspaces are *incident* if they intersect in such an $(n-1)$-subspace, in which case the two n-subspaces have opposite type. Again, the totally singular $(n-1)$-subspaces do not appear among the n types of objects of our geometry, because they are already represented as incident pairs of totally singular n-subspaces.

For objects of dimension $k \neq k'$ in our geometry, incidence is inclusion just as in the case of projective spaces. Given a totally singular k-subspace U, the **residue** of U is the subgeometry formed by the objects incident with U. This geometry has diagram obtained from the D_n diagram by deleting the corresponding node. Consider first the case $k \leq n-2$, in which the resulting diagram is of type $A_{k-1} \oplus D_{n-k}$. The objects of this residual geometry are the subspaces of U, forming the projective space $\mathbb{P}U \simeq \mathbb{P}^{k-1}(\mathbb{F}_q)$; *and* the totally singular subspaces containing U. The latter correspond to the totally singular subspaces of $U^\perp/U \simeq O^+_{2(n-k)}(q)$, these forming a geometry of type D_{n-k}. Furthermore every object of the A_{k-1}-geometry is incident with every member of the D_{n-k}-geometry, as is represented by the fact that there are no edges joining the corresponding subgraphs. In the case $k = n$, the residue of a totally singular n-space U form a geometry of type A_{n-1}, as we should expect since this is the diagram obtained from the D_n diagram by removing the node corresponding to U. In fact the objects of the residual geometry are the subspaces of U of dimension 1, 2, ..., $n-2$; *and* the totally singular n-subspaces $U' < V$ incident with U. As we have said, such subspaces U' are completely determined by the intersections $U' \cap U$, these being all the hyperplanes of U; so the residue of U is clearly isomorphic to $\mathbb{P}U \simeq \mathbb{P}^{n-1}(\mathbb{F}_q)$ as expected.

We devote Section 25 to the study of the important special case of the $O_6^+(q)$-quadric (the *Klein quadric*). The other particularly fascinating special case is that of $O_8^+(q)$-quadric, the *triality quadric,* so-called because it admits a symmetry of order 3 mapping

$$\begin{matrix}\text{singular}\\ \text{points}\end{matrix} \longrightarrow \begin{matrix}\text{totally singular}\\ \text{solids of type I}\end{matrix} \longrightarrow \begin{matrix}\text{totally singular}\\ \text{solids of type II}\end{matrix} \longrightarrow \begin{matrix}\text{singular}\\ \text{points}\end{matrix}$$

while mapping totally singular lines to totally singular lines. Regrettably, we will probably not have time to explain how such an automorphism is defined. Unlike the duality automorphism of projective spaces, which exists in every dimension, the triality automorphism is unique to the $O_8^+(q)$-quadric. Together with the automorphisms interchanging the two types of totally singular solids, we see that the automorphism group of the geometry induces the full symmetry group S_3 of the D_4 diagram.

24.10 Finite Classical Polar Spaces. We describe three types of finite classical polar spaces embedded in a finite-dimensional vector space V over $F = \mathbb{F}_q$. A **polarity** of $\mathbb{P}V$ is a map '$\perp$' of order 2 (so that $(X^\perp)^\perp = X$) interchanging points with hyperplanes, lines

with subspaces of codimension 2, etc., while preserving (or rather, reversing) incidence: $X_1 \subset X_2$ iff $X_2^\perp \subset X_1^\perp$. A subspace $X \leq V$ is **absolute** if $X \leq X^\perp$. All objects of the associated polar space, are nonzero subspaces of V which are absolute with respect to the associated polarity. The three types of polarities (orthogonal, symplectic and unitary) give rise to the three types of polar spaces. Although the orthogonal polar spaces are the most complicated of the three to define, we describe them first since most people find them closer to their experience.

- **Orthogonal Polar Spaces.** Let $Q : V \to F$ be a quadratic form, with associated bilinear form B; thus for all $x, y \in V$ and $a, a' \in F$,

  $$Q(x+y) = Q(x) + Q(y) + B(x,y);$$
  $$B(ax+a'x', y) = aB(x,y) + a'B(x',y).$$

  We require the form Q to be nondegenerate; the meaning of this requirement is recalled below. The **orthogonal polarity** is defined by

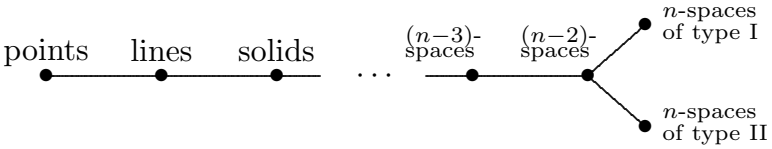
  $$X^\perp = \{v \in V : B(x,v) = 0 \text{ for all } x \in X\}, \text{ where } X \leq V.$$

  Consider first the case that q is odd. Here we require that B be nondegenerate, i.e. $V^\perp = 0$. The objects of the polar space are the nonzero subspaces $X \leq V$ such that $X \leq X^\perp$. These are just the nonzero subspaces lying in the associated quadric.

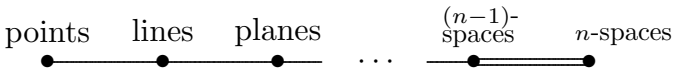
  For general q we require only that Q be nondegenerate, which means that $V^\perp$ contains no singular points. The objects of the polar space are the (nonzero) totally singular subspaces $X \leq V$, i.e. those which satisfy $Q(x) = 0$ for all $x \in X$. These are simply the nonzero subspaces contained in the associated quadric. (For q odd this description is equivalent to that given in the preceding paragraph. But for q even, the bilinear form B is alternating, so the nonzero subspaces satisfying $X \leq X^\perp$ give the symplectic polar space described below. In particular for q even, the absolute points with respect to B include *all* points of $\mathbb{P}V$, not just the points of a quadric.)

  In most cases, we define two objects of the polar space to be *incident* if, as subspaces of V, one properly contains the other. In the case of $O^+_{2n}(q)$-quadrics, however (see Section 24.9), we make an exception: two totally singular subspaces of dimension $\frac{n}{2}$ are *incident* iff they intersect in a subspace of dimension $\frac{n}{2}-1$. Moreover in this case, totally singular subspaces of dimension $\frac{n}{2}-1$ are not properly considered as objects of the polar space; instead they are identified as incident pairs of totally singular n-spaces.

  We have orthogonal polar spaces of hyperbolic type $O^+_{2n}(q)$, of parabolic type $O_{2n+1}(q)$, and of elliptic type $O^-_{2n}(q)$, according to the type of the corresponding quadratic form. Orthogonal polar spaces of type $O^+_{2n}(q)$ belong to the Coxeter-Dynkin diagram of type D_n:



as described above. Polar spaces of type $O_{2n+1}(q)$ belong to the Coxeter-Dynkin diagram of type B_n or C_n:



(The distinction between the Coxeter-Dynkin diagrams of type B_n and C_n arises when considering root systems and Lie algebras, but is not an issue here.) The meaning of the double bond between totally singular $(n-1)$-spaces and totally singular n-spaces, will be explained in Section 28. Polar spaces of type $O^-_{2n}(q)$ belong to the Coxeter-Dynkin diagram of type B_{n-1}:

An invertible linear transformation $g \in GL(V)$ is a **(linear) isometry** of the quadratic form Q if $Q(x^g) = Q(x)$ for all $x \in V$. The set of all such isometries is the **orthogonal group**, denoted $O^+_{2n}(q)$, $O_{2n+1}(q)$ or $O^-_{2n}(q)$ according to the type of the quadratic form, and the field order q. This group acts on the polar space. But not necessarily faithfully; for q odd, the normal subgroup $\{\pm I\}$ of order 2 acts trivially. Moreover in the $O^+_{2n}(q)$ case, only half the isometries are type-preserving; the other half interchange the two types of maximal totally singular subspaces.

- **Symplectic Polar Spaces.** Let $B : V \times V \to F$ be an alternating bilinear form; thus for all $x, x', y \in V$ and $a, a' \in F$,

  $B(ax+a'x', y) = aB(x, y) + a'B(x', y)$;
  $B(x, x) = 0$ and $B(y, x) = -B(x, y)$.

  For every subspace $X \leq V$ we define

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

  We require that B be *nondegenerate*; that is, $V^\perp = 0$. This condition requires that $\dim V = 2n$ be even, as follows from Exercise #22.1 by considering a matrix defining B. We refer to such a nondegenerate alternating bilinear form as simply a **symplectic form**. There is a version of Witt's Theorem 23.7 for symplectic forms. This has numerous consequences similar to Corollary 23.8; for example, all maximal totally isotropic subspaces have the same dimension (in this case, n). The **symplectic polarity** is the map $X \mapsto X^\perp$. The **symplectic polar space** $Sp_n(q)$ has as its objects the (nonzero) totally isotropic subspaces with respect to B, i.e. those nonzero subspaces $X \leq V$ such that $X \leq X^\perp$. This polar space belongs to the Coxeter-Dynkin diagram of type B_n (shown above).

  An invertible linear transformation $g \in GL(V)$ is a **(linear) isometry** of the alternating form B if $B(x^g, y^g) = B(x, y)$ for all $x, y \in V$. The set of all such isometries

is the **symplectic group** $Sp_n(q)$. This group acts on the symplectic polar space. But not necessarily faithfully; for q odd, the normal subgroup $\{\pm I\}$ of order 2 acts trivially.

- **Unitary Polar Spaces.** Here we must assume the field order $q = q_0^2$. The quadratic extension $\mathbb{F}_q \supset \mathbb{F}_{q_0}$ has an automorphism of order two given by $x \mapsto x^{q_0}$. Let $H : V \times V \to F$ be a **Hermitian form**; thus for all $x, x', y \in V$ and $a, a' \in F$,

  $H(ax+a'x', y) = aH(x, y) + a'H(x', y)$;
  $H(y, x) = H(x, y)^{q_0}$.

  For every subspace $X \leq V$ we define

  $X^\perp = \{v \in V \,:\, H(x, v) = 0 \text{ for all } x \in X\}$.

  Note that H is linear in its first argument, but semilinear (conjugate linear) in its second argument:

  $H(x, ay+a'y') = a^{q_0}H(x, y) + (a')^{q_0}H(x, y')$;

  so we say H is **sesquilinear**, i.e. $1\frac{1}{2}$-linear. We require that H be *nondegenerate*; that is, $V^\perp = 0$. There is also a version of Witt's Theorem 23.7 for Hermitian forms. So once again, all maximal totally isotropic subspaces have the same dimension, namely $n = \lfloor \frac{m}{2} \rfloor$ where $m = \dim V$. The **unitary polarity** is the map $X \mapsto X^\perp$. The **unitary polar space** $U_n(q_0)$ has as its objects those nonzero subspaces $X \leq V$ such that $X \leq X^\perp$. This polar space belongs to the Coxeter-Dynkin diagram of type B_n (shown above).

  An invertible linear transformation $g \in GL(V)$ is a **(linear) isometry** of the Hermitian form H if $H(x^g, y^g) = H(x, y)$ for all $x, y \in V$. The set of all such isometries is the **unitary group** $U_n(q_0)$. This group acts on the unitary polar space. The normal subgroup $\{\lambda I : \lambda^{q_0+1}=1\}$ of order q_0+1 acts trivially.

In each case if one wants the full automorphism group of the given polar space, one must consider semilinear maps rather than just linear maps. Moreover one must consider not only isometries (which actually preserve the relevant form) but **similarities**, which scale the form by a nonzero constant; thus for example in the orthogonal case $Q(x^g) = c_g Q(x)$ for all $x \in V$, where $c_g \in \mathbb{F}_q^\times$ may depend on g but not on x.

Polar spaces, like projective spaces, may be defined axiomatically using just points and lines; and just as one finds with projective spaces (Section 20), with exceptions only in small dimensions, all finite polar spaces turn out to be the classical ones listed above. We feel the beginning student should study the classical examples before worrying about to the axiomatic foundations.

Exercises 24.

1. Let V be the vector 4-space over a field F, and let $Q : V \to F$ be the determinant map.
   (a) Show that Q is a quadratic form.
   (b) Nonzero elements of V have rank 1 or 2, and elements of $\mathbb{P}V$ are distinguished as points of rank 1 and rank 2 accordingly. Show that the quadric defined by Q consists of all points of rank 1.

(c) Show that the corresponding quadric is hyperbolic, and that one regulus has lines parameterized by the (1-dimensional) row space of the underlying matrix; and the alternate regulus has lines parameterized by the (1-dimensional) column space of the underlying matrix.

2. Extend Theorem 24.6 by counting also the number of totally singular lines in each case (hyperbolic, parabolic and elliptic).

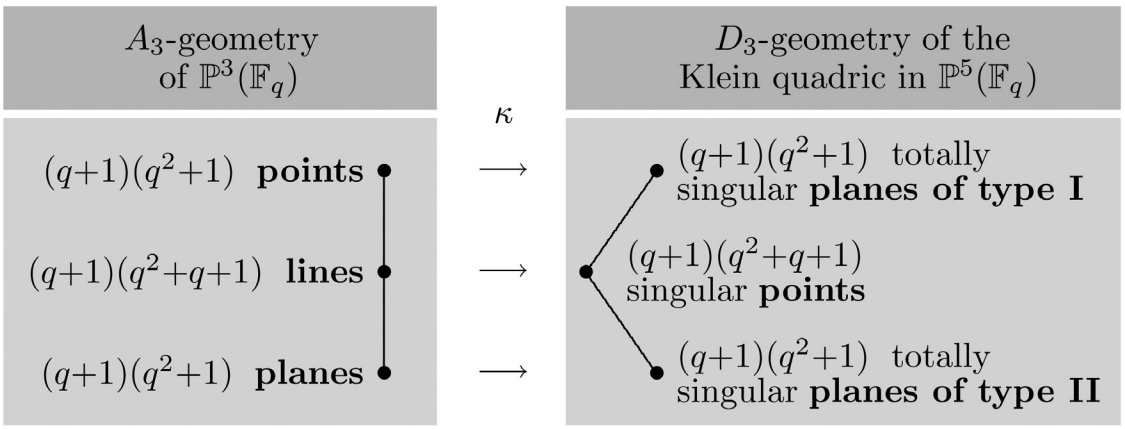
25. The Klein Correspondence

Let F be a field. Recall that every line ℓ in $\mathbb{P}^3(F)$ is specified by a 6-tuple of Plücker coordinates $(\ell_{01}, \ell_{02}, \ell_{03}, \ell_{12}, \ell_{13}, \ell_{23})$ which is determined by ℓ up to nonzero scalar multiple. The Plücker coordinates satisfy

$$\ell_{01}\ell_{23} - \ell_{02}\ell_{13} + \ell_{03}\ell_{02} = 0 \tag{25.1}$$

and the Plücker map $\ell \mapsto \langle(\ell_{01}, \ell_{02}, \ell_{03}, \ell_{12}, \ell_{13}, \ell_{23})\rangle$ is a bijection between the set of lines of $\mathbb{P}^3(F)$ and the point set of the **Klein quadric**, which is the quadric in $\mathbb{P}^5(F)$ defined by (25.1). Note that for a finite field $F = \mathbb{F}_q$, the Klein quadric is simply the *hyperbolic quadric*, otherwise known as the $O_6^+(q)$-quadric. Every line of the Klein quadric lies in exactly two planes of the quadric, one of each type. Two distinct planes of the quadric (necessarily of opposite type) are *incident* if they intersect in a line of the quadric. Two distinct planes which are not incident must be disjoint (if they are of opposite type) or intersect in a point (if they are of the same type).

The Plücker map described above extends to an isomorphism from the entire A_3-geometry of projective 3-space, to the D_3-geometry of the Klein quadric, as indicated by the following diagram. (Although this works for arbitrary fields, we indicate here the number of objects of each type in the corresponding geometry according to our previous counts. It is comforting to see that these numbers agree!)



This isomorphism of geometries, denoted here by κ, is known as the **Klein correspondence**. Its explicit description is conveniently described in the language of exterior algebra.

Let V be a vector 4-space over F, and consider the exterior algebra of V which is the 16-dimensional algebra

$$\bigwedge V = \bigoplus_{k\geq 0} \bigwedge^k V = F \oplus V \oplus \bigwedge^2 V \oplus \bigwedge^3 V \oplus \bigwedge^4 V$$

defined as in Appendix A4; here the summands have dimension 1, 4, 6, 4, 1 respectively. We take $\{e_0, e_1, e_2, e_3\}$ as a basis for V, so that

$\bigwedge^2 V$ has basis $\{e_0\wedge e_1,\ e_0\wedge e_2,\ e_0\wedge e_3,\ e_1\wedge e_2,\ e_1\wedge e_3,\ e_2\wedge e_3\}$;

$\bigwedge^3 V$ has basis $\{e_0\wedge e_1\wedge e_2,\ e_0\wedge e_1\wedge e_3,\ e_0\wedge e_2\wedge e_3,\ e_1\wedge e_2\wedge e_3\}$; and

$\bigwedge^4 V$ has basis $\{e_0\wedge e_1\wedge e_2\wedge e_3\}$.

The points, lines and planes of $\mathbb{P}V$ are mapped to points of $\bigwedge^k V$ for $k = 1, 2, 3$ respectively via the Plücker map $U \mapsto \bigwedge^k U$. In the case $k = 2$ this map is not surjective; its image is the Klein quadric which we denote $\mathcal{K} \subset \mathbb{P}\big(\bigwedge^2 V\big)$. This quadric is defined by the quadratic form

$$Q(v) = v_{01}v_{23} - v_{02}v_{13} + v_{03}v_{12} \quad \text{where } v = \sum_{0\leq i<j\leq 3} v_{ij}\, e_i\wedge e_j \in \bigwedge^2 V.$$

The associated bilinear form $B(u, v) = Q(u + v) - Q(u) - Q(v)$ satisfies

$$u \wedge v = B(u, v)\, e_0\wedge e_1\wedge e_2\wedge e_3 \quad \text{where } u, v \in \bigwedge^2 V.$$

We are now ready to define the Klein correspondence.

The Klein Correspondence κ

For a point $\langle x\rangle < V$, define

$$\kappa(\langle x\rangle) = x\wedge V = \big\{\langle v\rangle \in \bigwedge^2 V \,:\, x\wedge v = 0\big\}.$$

For a line $\ell = \langle x, y\rangle < V$, define

$$\kappa(\ell) = \ell\wedge\ell = \bigwedge^2 \ell = \langle x\wedge y\rangle \in \mathcal{K}.$$

For a plane $U = \langle x, y, z\rangle < V$, define

$$\kappa(U) = \bigwedge^2 U = \langle x\wedge y,\ x\wedge z,\ y\wedge z\rangle.$$

We proceed to show that κ has the required properties. For any line $\ell = \langle x, y\rangle < V$ where $x = \sum_{0\leq i\leq 3} x_i e_i$ and $y = \sum_{0\leq i\leq 3} y_i e_i$, we have

$$\kappa(\ell) = \kappa(\langle x, y\rangle) = \langle v\rangle \qquad \text{where } v = \sum_{0\leq i<j\leq 3} \begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} e_i\wedge e_j$$

which agrees with the image $\rho(\ell)$ under the Plücker map as defined previously. By Section 22, κ maps lines of $\mathbb{P}V$ bijectively to points of the Klein quadric $\mathcal{K}$. Also by Section 22 and remarks from Appendix A4 identifying $\bigwedge^2 V$ with skew-symmetric 4×4 matrices, we see that points of the Klein quadric are those represented as 'pure' wedge products $\langle x \wedge y \rangle$ for some linearly independent $x, y \in V$. The following observation is useful.

25.1 Lemma. Let ℓ and ℓ' be two lines of V. Then the following four conditions are equivalent.

(i) ℓ and ℓ' meet in a point of V.

(ii) ℓ and ℓ' lie in a plane of V.

(iii) The points $\kappa(\ell), \kappa(\ell') \in \mathcal{K}$ lie on a line of $\mathcal{K}$.

(iv) The points $\kappa(\ell), \kappa(\ell') \in \mathcal{K}$ are perpendicular relative to B.

Proof. Write $\ell = \langle x, y \rangle$, $\ell' = \langle x', y' \rangle$. Both (i) and (ii) are equivalent to the condition that x, y, x', y' are linearly dependent; this in turn is equivalent to the condition that

$$x \wedge y \wedge x' \wedge y' = 0$$

by (A4.5). Since $x \wedge y \wedge x' \wedge y' = B(x \wedge y,\, x' \wedge y')\, e_0 \wedge e_1 \wedge e_2 \wedge e_3$, the latter condition is equivalent to the points $\langle x \wedge y \rangle, \langle x' \wedge y' \rangle \in \mathcal{K}$ being perpendicular; this is condition (iv). Finally (iii) is equivalent to (iv) since every line of $\mathcal{K}$ is totally singular and hence totally isotropic; conversely any two singular points which are perpendicular, span a totally singular line. □

We are ready to verify that the Klein correspondence κ is an isomorphism from the A_3-geometry of $\mathbb{P}V$ to the D_3-geometry of $\mathcal{K}$. A point $\langle x \rangle$ of V lies in a line $\ell = \langle y, y' \rangle$ of V, iff $\{x, y, y'\}$ is linearly dependent, iff $x \wedge y \wedge y' = 0$, iff $\kappa(\ell) = \langle y \wedge y' \rangle$ lies in $\kappa(\langle x \rangle)$.

Let $\ell, U < V$ be a line and a plane respectively. If $\ell \subset U$ then ℓ meets every line $\ell' \subset U$, so the singular point $\kappa(\ell)$ is perpendicular to every point $\kappa(\ell')$ of $\kappa(U)$; thus $\langle \kappa(\ell), \kappa(U) \rangle$ is totally singular. But $\kappa(U)$ is a maximal totally singular subspace so the point $\kappa(\ell)$ must lie in $\kappa(U)$. Conversely if $\kappa(\ell)$ lies in the totally singular subspace $\kappa(U)$ then $\kappa(\ell)$ meets every point $\kappa(\ell')$ of $\kappa(U)$, so ℓ meets every line $\ell' \subset U$, so $\ell \subset U$.

Let $\langle x \rangle, U < V$ be a point and a plane of V respectively. If $\langle x \rangle$ lies in U then there exist lines $\ell, \ell' < V$ such that $\ell \cap \ell' = \langle x \rangle$ and $\langle \ell, \ell' \rangle = U$. Then $\kappa(\langle x \rangle) \cap \kappa(U)$ contains the totally singular line in $\mathcal{K}$ spanned by $\kappa(\ell)$ and $\kappa(\ell')$, so $\kappa(\langle x \rangle)$ is incident with $\kappa(U)$. Conversely, if $\kappa(\langle x \rangle)$ and $\kappa(U)$ are incident then $\kappa(\langle x \rangle) \cap \kappa(U)$ contains a singular point, which we know has the form $\kappa(\ell)$ for some line $\ell < V$ such that $\langle x \rangle \subset \ell \subset U$, so $\langle x \rangle$ is incident with U.

The Klein correspondence is immensely useful in understanding many features of the geometry of projective 3-space. For example recall that a spread of $\mathbb{P}^3(\mathbb{F}_q)$ is a collection of $q^2{+}1$ lines which partition the $(q^2{+}1)(q{+}1)$ points. This is equivalent to a set of $q^2{+}1$ mutually skew lines of $\mathbb{P}^3(\mathbb{F}_q)$; and by Lemma 25.1, such a set of lines is equivalent to a set of $q^2{+}1$ points on the Klein quadric, no two of which are perpendicular; i.e. no two on a line of the quadric. Such a point set is called an *ovoid* of the Klein quadric.

25.2 Example: Ovoids in $O_6^+(3)$. The quadratic form on $\mathbb{F}_3^6$ defined by $Q(v) = v_1^2 + v_2^2{+}\cdots{+}v_5^2{-}v_6^2$ has discriminant -1, and since $(-1)^3 \operatorname{disc} Q = 1$ is a square, Q is hyperbolic. (Note that the form $v_1^2 + v_2^2 + \cdots + v_6^2$ is elliptic over $\mathbb{F}_3$ and therefore unsuitable for our present purpose.) The associated bilinear form is

$$B(u,v) = Q(u+v) - Q(u) - Q(v) = 2(u_1v_1 + u_2v_2 + \cdots + u_5v_5 - u_6v_6)\,.$$

Consider the set $\mathcal{O}$ consisting of the $\binom{5}{3} = 10$ points of the form $\langle(1^3\,0^2|0)\rangle$; i.e. the last coordinate is zero, and the first five coordinates include three 1's and two 0's. Note that $\mathcal{O}$ consists of singular points, no two of which are perpendicular since 1's in distinct points overlap in either 1 or 2 positions. Also $|\mathcal{O}| = 3^3{+}1$ so $\mathcal{O}$ is an ovoid.

Likewise the set $\mathcal{O}'$ consisting of $2{\times}5 = 10$ points of the form $\langle(1^4\,0|\pm1)\rangle$; here the last coordinate is ±1 and the first five coordinates include four 1's and one 0. The point set $\mathcal{O}'$ is also an ovoid; for example

$$\begin{aligned}
B\big((111101),(111102)\big) &= 2(1+1+1+1+0-2) = 1 \neq 0;\\
B\big((111101),(111011)\big) &= 2(1+1+1+0+0-1) = 1 \neq 0;\\
B\big((111101),(111012)\big) &= 2(1+1+1+0+0-2) = 2 \neq 0.
\end{aligned}$$

By the Klein correspondence, both $\mathcal{O}$ and $\mathcal{O}'$ give rise to spreads of $\mathbb{P}^3(\mathbb{F}_3)$, and therefore translation planes of order $3^2 = 9$. These two planes are non-isomorphic, as we might expect from the fact that $\mathcal{O}$ spans only a 4-dimensional subspace of codimension 2, whereas $\mathcal{O}'$ spans the entire $O_6^+(3)$-space. In fact the translation planes associated to $\mathcal{O}$ and $\mathcal{O}'$ are the classical plane and the Hall plane of order 9, respectively.

25.3 Example: Ovoids from Slope Matrices [33]**.** Consider a set of $2{\times}2$ slope matrices for a spread over $\mathbb{F}_q$, say

$$M_i = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix}, \qquad i = 0,1,2,\ldots,q^2{-}1$$

where $M_i{-}M_j$ is nonsingular whenever $i \neq j$. The associated spread has components

$$\ell_i = \langle(1,0,a_i,b_i),(0,1,c_i,d_i)\rangle, \quad \ell_\infty = \langle(0,0,1,0),(0,0,0,1)\rangle.$$

By the Klein correspondence we obtain the points of the associated ovoid as

$$\begin{aligned}\langle(1,0,a_i,b_i)\wedge(0,1,c_i,d_i)\rangle &= \langle e_{01}{+}c_ie_{02}{+}d_ie_{03}{-}a_ie_{12}{-}b_ie_{13}{+}(a_ib_i{-}b_ic_i)e_{23}\rangle\\ &= \langle(1,\, c_i,\, d_i,\, -a_i,\, -b_i,\, a_ib_i{-}b_ic_i)\rangle;\\ \langle(0,0,1,0)\wedge(0,0,0,1)\rangle &= \langle e_{23}\rangle = \langle(0,0,0,0,0,1)\rangle\end{aligned}$$

with respect to the hyperbolic quadratic form

$$Q(v_1,v_2,\ldots,v_6) = v_1v_6 - v_2v_5 + v_3v_4\,.$$

For example if one takes the explicit spread matrices for the Hall plane of order 9, as given in Example 2.5, and constructs the associated ovoid as shown above, one may change coordinates so as to obtain the quadratic form of Example 25.2, in such a way that our ovoid gives the ovoid denoted $\mathcal{O}'$ in Example 25.2.

25.4 Example: Classical (regular) ovoids in $O_6^+(q)$. We may decompose any $O_6^+(q)$-space as $O_4^-(q) \perp O_2^-(q)$; simply take any elliptic line and its perp. By Theorem 24.6, the $O_4^-(q)$-subspace contains $q^2{+}1$ singular points; and these must form an ovoid, since an $O_4^-(q)$-space contains no totally singular lines. We show that these ovoids do indeed give rise to the classical translation planes of order q^2.

As explained in Section 3, the classical planes of order q^2 may be constructed as follows. Let $t^2+t+c \in \mathbb{F}_q[t]$ be an irreducible quadratic polynomial over $\mathbb{F}_q$. A regular spread is defined by the slope matrices

$$\begin{bmatrix} s & t \\ -ct & s{-}t \end{bmatrix}, \quad s,t \in \mathbb{F}_q$$

since these form a matrix representation of $\mathbb{F}_{q^2}$ as described in Appendix A1.2. The corresponding ovoid $\mathcal{O}$, constructed as in the previous example, consists of the points

$$\langle(0,0,0,0,0,1)\rangle; \qquad \langle(1,-ct,s{-}t,-s,t,s^2{-}st{+}ct^2)\rangle \text{ for } s,t\in\mathbb{F}_q$$

lying in $\langle x,y\rangle^\perp$ where $x=(0,1,-1,-1,0,0)$ and $y=(0,c,0,0,-1,0)$. Since

$$Q(ax+by) = Q(0,a{+}bc,-a,-a,-b,0) = a^2{+}ab{+}cb^2$$

we have $\langle x,y\rangle \simeq O_2^-(q)$ and $\mathcal{O} \subset \langle x,y\rangle^\perp \simeq O_4^-(q)$.

25.5 Example: $p \equiv 1 \mod 4$. The quadratic form $Q(v) = v_1^2+v_2^2+\cdots+v_6^2$ has associated bilinear form $B(u,v) = 2(u_1v_1+u_2v_2+\cdots+u_6v_6)$ and discriminant 2^6. In order for Q to be hyperbolic, we require that $(-1)^3 \operatorname{disc} Q = -64$ be a nonzero square in

$\mathbb{F}_q$, i.e. $q \equiv 1 \mod 4$. The following construction works only for fields $\mathbb{F}_p$ of prime order $p \equiv 1 \mod 4$. Consider the set of all 6-tuples $(a_1, a_2, \ldots, a_6)$ of *integers* such that

$$a_1^2 + a_2^2 + \cdots + a_6^2 = 6p; \qquad a_1 \equiv a_2 \equiv \cdots \equiv a_6 \equiv 1 \mod 4.$$

In [44] it is shown that there are exactly $p^2{+}1$ such vectors; and that they give an ovoid $\mathcal{O}$ in $O_6^+(p)$.

For $p = 5$ the ovoid $\mathcal{O}$ has 6 points of the shape $\langle(5^1, 1^5)\rangle$ (i.e. one 5 and five 1's) and $\binom{6}{3} = 20$ points of shape $\langle(-3^3, 1^3)\rangle$, for a total of $26 = 5^2{+}1$ points. After reducing mod 5, these points have shape $\langle(0, 1^5)\rangle$ and $\langle(1^3, 2^3)\rangle$. We check that these points form an ovoid:

$$\begin{aligned}
B\big((011111), (101111)\big) &= 2(0{+}0{+}1{+}1{+}1{+}1) \equiv 3 \not\equiv 0 \mod 5;\\
B\big((011111), (111222)\big) &= 2(0{+}1{+}1{+}2{+}2{+}2) \equiv 1 \not\equiv 0 \mod 5;\\
B\big((011111), (211122)\big) &= 2(0{+}1{+}1{+}1{+}2{+}2) \equiv 4 \not\equiv 0 \mod 5;\\
B\big((111222), (211122)\big) &= 2(2{+}1{+}1{+}2{+}4{+}4) \equiv 3 \not\equiv 0 \mod 5.
\end{aligned}$$

Similarly for $p = 13$ the ovoid $\mathcal{O}$ has

$$\begin{aligned}
&20 \text{ points of type } \langle(5^3, 1^3)\rangle;\\
&30 \text{ points of type } \langle(-7, 5, 1^4)\rangle;\\
&60 \text{ points of type } \langle(5^2, -3^3, 1^2)\rangle; \text{ and}\\
&60 \text{ points of type } \langle(-7, -3^3, 1^2)\rangle
\end{aligned}$$

for a total of $170 = 13^2{+}1$ points.

Exercises 25.

1. Construct an explicit ovoid in $O_6^+(17)$ by the method of Example 25.5.

26. Ovoids and Spreads of Projective Space

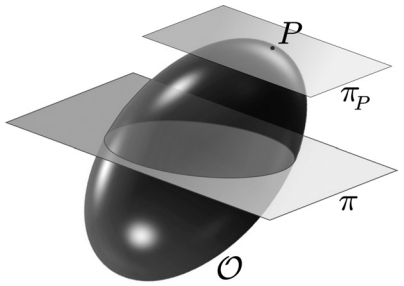
Here we consider the terms '*ovoid*' and '*spread*' as they are used in the context of projective space. In Section 27 we will encounter distinct, albeit related, usages of these terms in the context of polar spaces, particularly quadrics; this will be explained later. We begin with ovoids of projective 3-space.

A k-**arc** of $\mathbb{P}^3(F)$, or simply an **arc**, is a set $\mathcal{O}$ of k points, no three of which are collinear.

26.1 Theorem. An arc in $\mathbb{P}^3(\mathbb{F}_q)$ for $q > 2$ has at most $q^2{+}1$ points.

For $q > 2$, a $(q^2{+}1)$-arc of $\mathbb{P}^3(\mathbb{F}_q)$ (the maximum possible size) is called an **ovoid** of the projective 3-space. The failure of the indicated bound in the case $q = 2$ is discussed in Exercise #1.

26.2 Theorem. Let $\mathcal{O}$ be an ovoid of $\mathbb{P}^3(\mathbb{F}_q)$, and let $P \in \mathcal{O}$. Then there is a unique plane π_P (the 'tangent plane' at P) such that $\pi_P \cap \mathcal{O} = \{P\}$. Every plane π meets $\mathcal{O}$ either in a single point (i.e. π is a tangent plane) or in an oval of π.



Proofs of Theorems 26.1 and 26.2 are included below, only in the case q is odd; for proofs in the general case see [55], [3]. First we describe the known examples of ovoids. For all q, an elliptic quadric in $\mathbb{P}^3(\mathbb{F}_q)$ is an ovoid. Indeed if $\mathcal{O}$ is an $O_4^-(q)$ quadric then $|\mathcal{O}| = q^2+1$ by Theorem 24.6; and if a line ℓ meets $\mathcal{O}$ in three points then ℓ is totally singular by Exercise #23.5, a contradiction. Barlotti [3] showed that if every oval arising as a plane section of $\mathcal{O}$ is a conic, then $\mathcal{O}$ is an elliptic quadric; by Segre's Theorem 12.14, it follows that for q odd, every ovoid is an elliptic quadric. For $q = 2^r$ there is another infinite family of ovoids known, the *Suzuki-Tits ovoids* which arise for odd exponent $r \geq 3$, i.e. $q \in \{8, 32, 128, 512, \ldots\}$. It is not known whether every ovoid in projective 3-space for q even must be either an elliptic quadric or a Suzuki-Tits ovoid. This question has stood for several decades as one of the reigning open problems in finite geometry. The most significant progress in this direction to date, is a theorem of Brown [8] showing that every ovoid having *at least one* conic as a plane section, is an elliptic quadric. All ovoids for $q \leq 32$ are known, using computer results which rely on the classification of ovals for these small values of q.

Proof of Theorems 26.1 and 26.2 for q odd. Let $\mathcal{O}$ be a k-arc in $\mathbb{P}^3(\mathbb{F}_q)$. We may assume $k \geq 2$, and let the points $P, Q \in \mathcal{O}$ be distinct. By the restriction on q, every plane meets $\mathcal{O}$ in at most $q+1$ points. There are exactly $q+1$ planes containing the line PQ, each of which contains at most $q-1$ points of $\mathcal{O}$ other than P and Q; this gives

$$|\mathcal{O}| \leq 2 + (q+1)(q-1) = q^2 + 1$$

which is the conclusion of Theorem 26.1. Henceforth we assume $k = |\mathcal{O}| = q^2 + 1$. We proceed using counting arguments similar to those arising in the proof of Proposition 18.1. Let m_k be the number of planes of $\mathbb{P}^3(\mathbb{F}_q)$ meeting $\mathcal{O}$ in exactly k points for $k \in \{0, 1, 2, \ldots, q+1\}$. The total number of planes is

$$\sum_{0 \leq k \leq q+1} m_k = \begin{bmatrix} 4 \\ 3 \end{bmatrix}_q = (q^2+1)(q+1). \tag{26.3}$$

Counting in two different ways the number of pairs (P, π) where π is a plane and $P \in \pi \cap \mathcal{O}$ gives

$$\sum_{0\le k\le q+1} km_k = (q^2+1)(q^2+q+1). \tag{26.4}$$

Here we have used the fact that each of the q^2+1 points of $\mathcal{O}$ lies in exactly q^2+q+1 planes. Next we count in two different ways the number of ordered triples (P,Q,π) with π any plane, and the points $P,Q\in\pi\cap\mathcal{O}$ distinct:

$$\sum_{0\le k\le q+1} k(k-1)m_k = q^2(q^2+1)(q+1). \tag{26.5}$$

Here we have used the fact that there are $q^2(q^2+1)$ ordered pairs of points (P,Q) in $\mathcal{O}$, and each such pair determines a line PQ which lies in exactly $q+1$ planes. Finally we count in two different ways the number of ordered 4-tuples (P,Q,R,π) with π any plane, and the points $P,Q,R\in\pi\cap\mathcal{O}$ distinct:

$$\sum_{0\le k\le q+1} k(k-1)(k-2)m_k = q^2(q^4-1). \tag{26.6}$$

Here we use the fact that $\mathcal{O}$ contains $q^2(q^2+1)(q^2-1)$ ordered triples (P,Q,R) of points, each of which determines a unique plane. From (26.5) and (26.6) we obtain

$$\begin{aligned}\sum_{0\le k\le q+1} k(k-1)(q+1-k)m_k &= (q-1)\sum_{0\le k\le q+1} k(k-1)m_k \;-\sum_{0\le k\le q+1} k(k-1)(k-2)m_k\\ &= (q-1)q^2(q^2+1)(q+1) - q^2(q^4-1)\\ &= 0.\end{aligned}$$

We must have $m_k=0$ for all $k\in\{2,3,\dots,q\}$; otherwise the left side in the preceding equality would be positive. Now the equations (26.3)–(26.5) become

$$\begin{bmatrix}1 & 1 & 1\\ 0 & 1 & q+1\\ 0 & 0 & q(q+1)\end{bmatrix}\begin{bmatrix}m_0\\ m_1\\ m_{q+1}\end{bmatrix} = \begin{bmatrix}(q^2+1)(q+1)\\ (q^2+1)(q^2+q+1)\\ q^2(q^4-1)\end{bmatrix}$$

which has the unique solution $(m_0,m_1,m_{q+1})=(0,q^2+1,q(q^2+1))$. □

26.7 Inversive Planes from Ovoids. Let $\mathcal{O}$ be an ovoid in $\mathbb{P}^3(\mathbb{F}_q)$. Recall that $\mathbb{P}^3(\mathbb{F}_q)$ has $\begin{bmatrix}4\\1\end{bmatrix}_q = (q^2+1)(q+1)$ planes, q^2+1 of which are tangent to $\mathcal{O}$; the remaining $q(q^2+1)$ planes meet $\mathcal{O}$ in ovals. It is not hard to see that the incidence structure $(\mathcal{O},\mathcal{C})$ formed by the q^2+1 points of $\mathcal{O}$ and the $q(q^2+1)$ ovals in $\mathcal{O}$, form an inversive plane of order q (see Exercise #2.3). An inversive plane arising in this way is called **egglike**. Note that this construction mimics the construction of the real inversive plane in Exercise #2.3(a).

A significant open problem in finite geometry is to classify finite inversive planes up to isomorphism. This question can be divided into two parts:

(I) Must every finite inversive plane be egglike (i.e. arise from an ovoid in a finite projective 3-space)?

(II) What are the ovoids in finite projective 3-space? (This will tell us what the egglike inversive planes are.)

In the direction of (I), it is still unknown whether every finite inversive plane has prime-power order. Since an inversive plane of order n yields an affine plane of order n by Exercise #2.3(b), and hence a projective plane of order n by Section 7, the prime power conjecture for finite projective planes would imply the prime power conjecture for finite inversive planes. But it may be possible to prove that every finite inversive plane has prime power order, without proving that every finite projective plane has prime power order is reversible; this is because the converse of Exercise #2.3(b) might not hold.

It *is known* [24] that every finite inversive plane of even order is egglike (in particular its order must be a power of 2), answering question (I) affirmatively in the even order case; but in this case question (II) is open, as we have explained. By contrast, question (II) is settled in the odd order case as we have said, but in this case question (I) is open! So the original question of classifying finite inversive planes remains open, both in the even order case and in the odd order case, but for different reasons!

Traditionally, ovoids are defined more generally in projective n-space; and then it is proved that they do not exist for $n > 3$! For $n = 2$ an ovoid is simply an oval. 'Nuff said about ovoids.

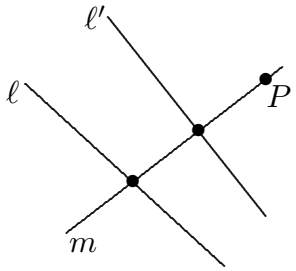
Recall that a *spread* of a $2k$-dimensional vector space V over F, is a collection S consisting of k-dimensional subspaces of V, which induce a partition of the nonzero vectors of V. We now interpret this definition projectively: A **spread** of $\mathbb{P}^{2k-1}(F)$ is a partition of the points into projective $(k-1)$-subspaces. In particular for $F = \mathbb{F}_q$ we must partition the $\begin{bmatrix} 2k \\ 1 \end{bmatrix}_q = (q^{2k}-1)/(q-1)$ points into subspaces each having $\begin{bmatrix} k \\ 1 \end{bmatrix}_q = (q^k-1)/(q-1)$ points, so the number of such spread members (i.e. components) is

$$\frac{(q^{2k}-1)/(q-1)}{(q^k-1)/(q-1)} \;=\; q^k{+}1$$

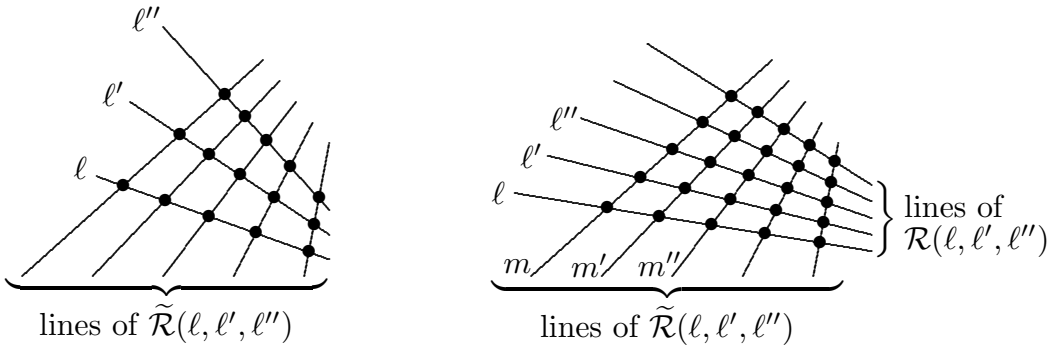
as we saw in Section 3. A **translation net** is a collection of mutually disjoint projective $(k-1)$-subspaces; thus a spread is a translation net having q^k+1 members, the maximum possible number. Recall that every field extension $E \supseteq F$ of degree k gives rise to a spread of $\mathbb{P}^{2k-1}(F)$ by taking as components all the one-dimensional E-subspaces of $E^2 = E \oplus E$, interpreted as F-subspaces of dimension k. Such spreads are called *regular*; and this term is justified by a geometric condition which posits the existence of large numbers of *reguli* in the spread. (For once the word 'regular', so over-used in mathematical contexts, is

etymologically motivated.) Rather than explain this here in full generality, we focus on the case of projective 3-space.

Let V be a 4-dimensional vector space over a field F. Consider any two skew (i.e. non-intersecting) lines ℓ, ℓ' in $\mathbb{P}V = \mathbb{P}^3(F)$. If P is any point not on $\ell \cup \ell'$ then there is a unique line m through P meeting both ℓ and ℓ'. This is not hard to see since $\langle P, \ell\rangle$ and $\langle P, \ell'\rangle$ are necessarily distinct planes, which therefore meet in the required line m. We call m the **transversal** to ℓ and ℓ' through P.



A collection of lines is **mutually**[1] **skew** if any two distinct lines in the collection are skew. Let ℓ, ℓ', ℓ'' be three mutually skew lines. Every point $P \in \ell''$ lies on a unique line meeting both ℓ and ℓ'; and as P varies over all points of ℓ'', this process generates the family $\widetilde{\mathcal{R}}(\ell, \ell', \ell'')$ of all lines **transversal** to each of the original three lines ℓ, ℓ', ℓ', i.e. meeting each of the lines ℓ, ℓ', ℓ''. Now by the preceding comments, the lines of $\widetilde{\mathcal{R}}(\ell, \ell', \ell'')$ are themselves mutually skew. Now let $m, m', m'' \in \widetilde{\mathcal{R}}(\ell, \ell', \ell'')$ be distinct; then $\widetilde{\mathcal{R}}(m, m', m'')$



is likewise a family of mutually skew lines containing the original three lines ℓ, ℓ', ℓ''; we denote

$$\mathcal{R}(\ell, \ell', \ell'') = \widetilde{\mathcal{R}}(m, m', m'') \tag{26.8}$$

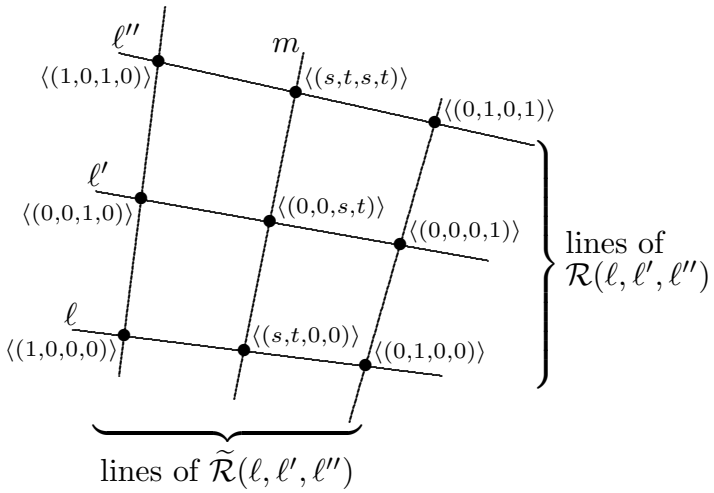
and we call this the **regulus** generated by the original three lines ℓ, ℓ' and ℓ''. The first family of mutually skew lines,

$$\mathcal{R}(m, m', m'') = \widetilde{\mathcal{R}}(\ell, \ell', \ell''), \tag{26.9}$$

is also a regulus; and the two reguli (26.8) and (26.9) are **alternate**, i.e. opposite to one another. We proceed to show that *every* member of one regulus is incident with every member of the alternate regulus; and that these two reguli are the ruling lines of a hyperbolic quadric in $\mathbb{P}V$.

[1] Some would say 'pairwise' skew; see the footnote on p.19.

We may coordinatize the lines of our reguli as follows: First choose bases $\{e_0, e_1\}$ for ℓ and $\{e_2, e_3\}$ for ℓ'. Now $\{e_0, e_1, e_2, e_3\}$ is a basis for V; and there is no loss of generality in assuming this to be the standard basis $\{(1,0,0,0), \ldots, (0,0,0,1)\}$. Evidently $\ell'' = \langle (1,0,a,b), (0,1,c,d)\rangle$ where the 2×2 matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ is non-singular. We may take $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, after performing a change of basis for ℓ' if necessary; thus $\ell'' = \langle (1,0,1,0), (0,1,0,1)\rangle$. The unique line of $\widetilde{\mathcal{R}}(\ell, \ell', \ell'')$ passing through a typical point $\langle (s,t,0,0)\rangle \in \ell$ is the line $m = \langle (s,t,0,0), (0,0,s,t)\rangle$ shown:



All points of m are singular with respect to the hyperbolic quadratic form

$$Q(x) = x_0 x_3 - x_1 x_2 \,.$$

The regulus $\widetilde{\mathcal{R}}(\ell, \ell', \ell'')$ consists of all lines of the form

$$\langle (s,t,0,0),\ (0,0,s,t)\rangle, \qquad (0,0) \neq (s,t) \in F^2$$

while the alternate regulus $\mathcal{R}(\ell, \ell', \ell'')$ consists of all lines of the form

$$\langle (s,0,0,t),\ (0,t,s,0)\rangle, \qquad (0,0) \neq (s,t) \in F^2.$$

All of the preceding discussion of reguli is greatly clarified and simplified using the Klein correspondence. Let ℓ, ℓ', ℓ'' be three mutually skew lines in $\mathbb{P}^3(F)$. The Klein correspondence maps these three lines to three mutually nonperpendicular points P, P', P'' of the $O_6^+(q)$-quadric. The plane $\pi = \langle P, P', P''\rangle$ spanned by these three points must therefore be nondegenerate, which means that π intersects the Klein quadric in a conic. The $q{+}1$ points of this conic, including P, P', P'', correspond (by the Klein correspondence to the regulus $\mathcal{R}(\ell, \ell', \ell'')$ generated by the three original lines. Moreover since $\pi \simeq O_3(q)$ is nondegenerate, the plane $\pi^\perp$ is also nondegenerate and disjoint from π; we have $F^6 = \pi \oplus \pi^\perp = \pi \perp \pi^\perp$. The plane $\pi^\perp$ intersects the Klein quadric in another conic, whose $q{+}1$ points correspond (by the Klein correspondence) to the alternate regulus $\widetilde{\mathcal{R}}(\ell, \ell', \ell'')$. All

of this follows easily from the known properties of the Klein correspondence, using the fact that every point of $\pi^\perp$ is perpendicular to every point of π.

A spread $\mathcal{S}$ of $\mathbb{P}^3(F)$ is **regular** if for all choices of three distinct lines $\ell, \ell', \ell'' \in \mathcal{S}$, we have the entire regulus $\mathcal{R}(\ell, \ell', \ell'') \subseteq \mathcal{S}$. By the Klein correspondence, this means that the corresponding ovoid $\mathcal{O}$ (justifiably called a **regular ovoid**) in the Klein quadric has the property that for any three of its points P, P', P'', all $q+1$ singular points of the plane $\langle P, P', P'' \rangle$ must also belong to $\mathcal{O}$. It is not hard to see that every such ovoid lies in an $O_4^-(q)$-subspace, and so the translation plane defined by any regular spread is isomorphic to $\mathbb{A}^2(\mathbb{F}_{q^2})$ (see Example 25.4).

Now let Σ be a spread of $\mathbb{P}^3(F)$. If Σ contains all the lines of some regulus, then clearly these lines may be replaced by the lines of the alternate regulus (since the alternate regulus covers the same points as the original regulus). This gives a new spread, and thereby a new translation plane which will typically not be isomorphic to the plane defined by Σ. For example the regular spread in $\mathbb{P}^3(\mathbb{F}_3)$ contains many reguli. Replacing one of these by its alternate regulus gives a spread which is not regular; this defines the Hall plane of order 9. To describe this process in terms of an ovoid $\mathcal{O}$ of the Klein quadric $\mathcal{K}$: If $\mathcal{O}$ contains a conic (which has the form $\pi \cap \mathcal{K}$ for some nondegenerate plane π) then by deleting this conic and replacing it by the conic in $\pi^\perp \cap \mathcal{K}$, we obtain a new ovoid.

Exercises 26.

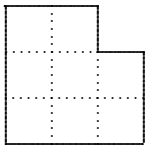
1. What is the maximum size of an arc in $\mathbb{P}^3(\mathbb{F}_2)$, and what is the structure of an arc attaining this maximum size?

   Hint. Consider the complement of a plane.

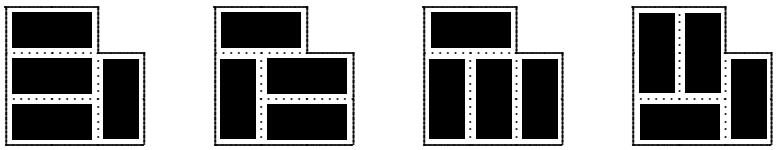
2. Let $\mathcal{C}$ be the binary code spanned by the planes of $\mathbb{P}^3(\mathbb{F}_q)$.
   (a) What is the dimension of $\mathcal{C}$? (Use Theorem 21.1.)
   (b) If $\mathcal{O}$ is an ovoid in $\mathbb{P}^3(\mathbb{F}_q)$, show that the collection of tangent planes $\{\pi_P : P \in \mathcal{O}\}$, in the notation of Theorem 26.2, is a linearly independent subset of $\mathcal{C}$.
   (b) Assuming that q is even, prove a stronger condition: that the tangent planes as in (b) actually form a basis for $\mathcal{C}$.

27. Ovoids and Spreads of Polar Spaces

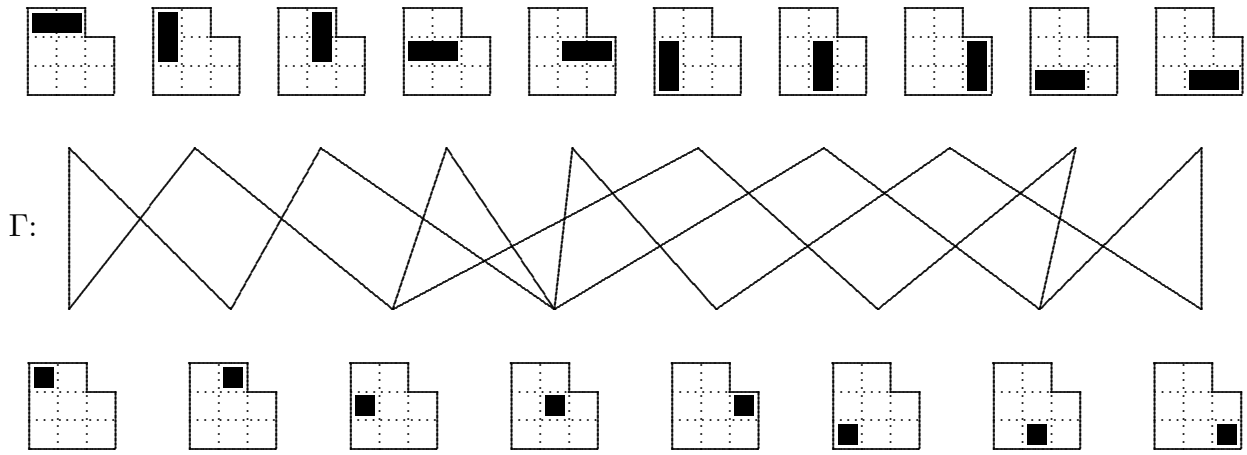
I personally think of ovoids and spreads of polar spaces as certain kinds of tilings or packings, and this is how I prefer to introduce them. Consider the problem of tiling the figure shown (which has been subdivided into eight unit squares) using 2×1 dominoes.



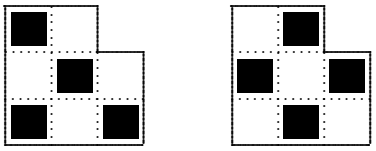
We require that the entire figure be covered, with no overlapping dominoes; so clearly 4 dominoes will be required. This problem has exactly four solutions, as shown:



This problem may be represented graphically as follows. Consider the bipartite graph Γ with 18 vertices corresponding to the 10 domino positions and the 8 cells of the figure. Edges in the graph Γ indicate which dominoes cover which cells, as shown:



The problem of tiling the figure with dominoes, is equivalent to finding a subset $\mathcal{S}$ of the top 10 vertices, so that each of the 8 bottom vertices is adjacent to exactly one member of $\mathcal{S}$ in the graph Γ. A related (or perhaps, dual) problem is that of finding a subset $\mathcal{O}$ of the bottom 8 vertices, such that each of the top 10 vertices is adjacent to a unique member of $\mathcal{O}$ in the graph Γ. This problem has two solutions:

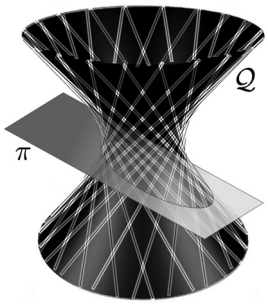


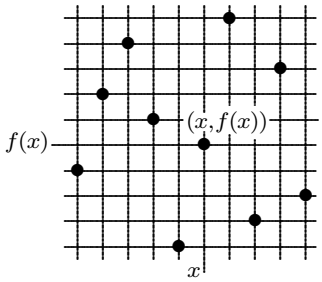
We refer a solution $\mathcal{S}$ of the tiling problem, as a *spread;* and a solution $\mathcal{O}$ of the dual problem, as an *ovoid.*

Likewise, given a polar space $\mathcal{P}$, we may consider the bipartite graph whose two types of vertices correspond to the points and maximal subspaces of $\mathcal{P}$. A **spread** (think: tiling) of the polar space is a subset of the set of maximal subspaces of $\mathcal{P}$, such that every point of $\mathcal{P}$ lies in a unique member of $\mathcal{S}$ (so that $\mathcal{S}$ partitions the points of the polar space). An **ovoid** of $\mathcal{P}$ is a collection of points of $\mathcal{P}$, such that every maximal subspace contains a unique point of $\mathcal{O}$. The ovoids of $O_6^+(q)$-quadrics introduced in Section 25 are in fact

examples of ovoids in our currently used sense of the term, as we shall see. First let us examine the situation for real quadrics.

The hyperbolic quadric $\mathcal{Q}$ in $\mathbb{P}^3(\mathbb{R})$ contains two spreads, these being the two reguli of the spread (the two families of lines ruling the quadric). It has *many* ovoids. For any plane π, either π is tangent to the quadric, in which case $\pi \cap \mathcal{Q}$ is a pair of intersecting lines; or $\pi \cap \mathcal{Q}$ is a conic meeting every line of $\mathcal{Q}$ in a unique point. In the latter case $\pi \cap \mathcal{Q}$ is an ovoid of $\mathcal{Q}$, and we obtain $2^{\aleph_0}$ ovoids of $\mathcal{Q}$ in this way. However there are many more ovoids: Since $\mathcal{Q}$ is a grid formed by two families of lines, each parameterized by $S^1 = \mathbb{P}^1(\mathbb{R}) = \mathbb{R} \cup \{\infty\}$, the point-line incidence structure of $\mathcal{Q}$ is isomorphic to $(\mathfrak{P}, \mathfrak{L})$ where $\mathfrak{P} = S^1 \times S^1 = \{(x, y) : x, y \in S^1\}$ and lines $\ell \in \mathfrak{L}$ are point sets of the form $x = \text{constant}$ or $y = \text{constant}$. Ovoids are point sets $\mathcal{O} \subset \mathfrak{P}$ of the form $\mathcal{O}_f = \{(x, f(x)) : x \in S^1\}$ where $f : S^1 \to S^1$ is any bijection. Since $|S^1| = 2^{\aleph_0}$, there are $2^{2^{\aleph_0}} = \beth_2$ such ovoids. The conics arise just for those bijections f expressible as fractional linear transformations.





The elliptic quadric in $\mathbb{P}^3(\mathbb{R})$ contains a unique ovoid and a unique spread; both are simply the entire point set of the quadric. Every (nonempty) nondegenerate quadric $\mathcal{Q}$ in $\mathbb{P}^{n-1}(\mathbb{R})$ is can be defined by an equation of the form

$$x_1^2 + \cdots + \cdots + x_m^2 - x_{m+1}^2 - \cdots - x_n^2 = 0$$

where $1 \le m \le \lfloor \frac{n}{2} \rfloor$; an example of an ovoid in $\mathcal{Q}$ is the set of points of $\mathcal{Q}$ lying in the subspace $\langle e_m, e_{m+1}, \ldots, e_n \rangle$ where $\{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{R}^n$.

Let us now determine the size of an ovoid or of a spread, for quadrics over finite fields.

27.1 Theorem. In any finite quadric, the size of an ovoid or of a spread is listed in the accompanying table. These values also give the maximum size of any set $\mathcal{S}$ of mutually disjoint maximal totally singular subspaces, and of any set $\mathcal{O}$ of mutually non-perpendicular singular points.

isometry type	$\lvert\mathcal{O}\rvert$	$\lvert\mathcal{S}\rvert$
$O^+_{2k}(q)$	$q^{k-1}+1$	$q^{k-1}+1$
$O^-_{2k}(q)$	q^k+1	q^k+1
$O_{2k+1}(q)$	q^k+1	q^k+1

It may at first come as a surprise that the size of an ovoid and of a spread coincide. An explanation for this fact is available using the language of m-systems; see [59] for details.

Proof of Theorem 27.1. Let $Q : V \to \mathbb{F}_q$ be a nondegenerate quadratic form. Every collection $\mathcal{S}$ of mutually disjoint maximal totally singular subspaces satisfies

$$|\mathcal{S}| \;\le\; \frac{|\{\text{sing. pts.}\}|}{|\{\text{pts. in each max. tot. sing. subspace}\}|}$$

$$= \begin{cases} \dfrac{(q^{k-1}+1)(q^k-1)/(q-1)}{(q^k-1)/(q-1)} \;=\; q^{k-1}+1, & \text{in case } O^+_{2k}(q); \\ \dfrac{(q^k+1)(q^{k-1}-1)/(q-1)}{(q^{k-1}-1)/(q-1)} \;=\; q^k+1, & \text{in case } O^-_{2k}(q); \\ \dfrac{(q^{2k}-1)/(q-1)}{(q^k-1)/(q-1)} \;=\; q^k+1, & \text{in case } O_{2k+1}(q) \end{cases}$$

and equality holds iff $\mathcal{S}$ is a spread.

Now let $\mathcal{O}$ be a collection of mutually nonperpendicular singular points; equivalently, $\mathcal{O}$ is a set of singular points, no two on a line of the quadric, and therefore no two in the same maximal totally singular subspace. By similar arguments as in the spread case (by considering the graph Γ described above) the upper bound for $|\mathcal{O}|$ occurs when every maximal totally singular subspace contains a unique point of $\mathcal{O}$. It therefore suffices to assume $\mathcal{O}$ is an ovoid.

For each isometry type of Q we count in two different ways the number of pairs $(\langle x\rangle, U)$ where U is a maximal subspace of the quadric containing a point $\langle x\rangle \in \mathcal{O}$, thus:

$$\begin{aligned} \big|\{U \le V \,:\, U \text{ max. tot. sing.}\}\big| &= |\mathcal{O}|\big|\{U \le V \,:\, U \text{ max. tot. sing., } \langle x\rangle \subseteq U\}\big| \\ &= |\mathcal{O}|\big|\{\text{max. tot. sing. subsp. of } x^\perp/\langle x\rangle\}\big| \\ &\quad \text{where } \langle x\rangle \text{ is a typical singular point.} \end{aligned} \tag{27.3}$$

If $Q : V \to F$ is of type $O^+_{2k}(q)$ then $x^\perp/\langle x\rangle \simeq O^+_{2(k-1)}(q)$ so by Table 24.6, (27.3) gives

$$2(q+1)(q^2+1)\cdots(q^{k-1}+1) = |\mathcal{O}| \cdot 2(q+1)(q^2+1)\cdots(q^{k-2}+1)$$

and so $|\mathcal{O}| = q^{k-1}+1$ as required. Similarly if Q is of type $O^-_{2k}(q)$ then $x^\perp/\langle x\rangle \simeq O^-_{2(k-1)}(q)$ and so

$$(q^2+1)(q^3+1)\cdots(q^k+1) = |\mathcal{O}| \cdot (q^2+1)(q^3+1)\cdots(q^{k-1}+1)$$

whence $|\mathcal{O}| = q^k+1$. Finally if Q is of type $O_{2k+1}(q)$ then $x^\perp/\langle x\rangle \simeq O^-_{2k-1}(q)$ and so

$$(q+1)(q^2+1)\cdots(q^k+1) = |\mathcal{O}| \cdot (q+1)(q^2+1)\cdots(q^{k-1}+1)$$

which also gives $|\mathcal{O}| = q^k+1$. □

27.4 Examples of ovoids. An ovoid in a quadric of type $O^+_2(q)$, $O_3(q)$ or $O^-_4(q)$ consists of the entire quadric in each case: two singular points, a conic, or an elliptic quadric

in $\mathbb{P}^3(\mathbb{F}_q)$ respectively. Ovoids in the Klein quadric $O_6^+(q)$ consist of q^2+1 mutually non-perpendicular points, and are equivalent to spreads of $\mathbb{P}^3(\mathbb{F}_q)$ via the Klein correspondence; examples of these were given in Section 25. Every $O_{2k}^-(q)$-quadric is embedded in an $O_{2k+1}(q)$-quadric as a hyperplane section, which in turn is similarly embedded in an $O_{2k+2}^+(q)$-quadric. Therefore any ovoid in $O_{2k}^-(q)$ is automatically an ovoid in $O_{2k+1}(q)$ and in $O_{2k}^+(q)$; an example of this principle gives the embedding of the $O_4^-(q)$-quadric as an ovoid in the Klein quadric $O_6^+(q)$. Similarly every ovoid $O_5(q)$ is automatically an ovoid in $O_6^+(q)$.

Consider the standard quadratic form $Q(x) = x_1^2 + \cdots + x_8^2$ which is nondegenerate for all odd q, defining an $O_8^+(q)$ quadric. In this case an ovoid $\mathcal{O}$ consists of q^3+1 singular points, no two of which are perpendicular with respect to the standard dot product. An example for $q=3$ is given by the set of $\binom{8}{2} = 28 = 3^3+1$ points of the form $\langle(0^2 1^6)\rangle$, i.e. points spanned by vectors with two 0's and six 1's. This example lies in the nondegenerate hyperplane $(1,1,\ldots,1)^\perp$ and so it is in fact an ovoid in an $O_7(3)$-quadric. An example for $q=5$ is given by:

$8 \times 7 = 56$ points of the shape $\langle(01^6 3)\rangle$;
$\binom{8}{4} = 70$ points of the shape $\langle(1^4 2^4)\rangle$

for a total of $126 = 5^3+1$ points. Similar constructions exist for every prime p. For $p=2$ we need $2^3+1 = 9$ points but the quadratic form must be replaced by a nondegenerate one.

To construct an ovoid in $O_8^+(2)$ (the unique ovoid in this space, up to isometry) we proceed as follows. Recall that the *weight* of a vector $x \in \mathbb{F}_2^9$, denoted $wt(x) \in \{0,1,2,\ldots,9\}$, is the number of nonzero coordinates. The subspace $V = (1,1,1,\ldots,1)^\perp < \mathbb{F}_2^9$ consists of all even weight vectors, i.e. vectors of weight $0, 2, 4, 6$ or 8. Define the quadratic form $Q : V \to \mathbb{F}_2$ by

$$Q(x) = \sum_{1 \le i < j \le 9} x_i x_j \,.$$

Note that if $wt(x) = k$, then $Q(x) \equiv \binom{k}{2} \mod 2$ so x is singular iff $k \in \{0,4,8\}$. In order to show that Q is nondegenerate, we show that for every nonzero vector $u \in V$ there exists $x \in V$ such that $Q(x+u) \neq Q(x)$, i.e. $B(u,x) \neq 0$. Since the argument depends only on $wt(u)$ it suffices to consider

$u = (111100000)$, $x = (100010000)$ so that $wt(x) = 2$, $wt(x+u) = 4$ so that $Q(x) = 1 \neq 0 = Q(x+u)$;

$u = (111111110)$, $x = (100000001)$ so that $wt(x) = 2$, $wt(x+u) = 8$ so that $Q(x) = 1 \neq 0 = Q(x+u)$.

Thus Q is nondegenerate. The number of singular points $\langle x\rangle$ is $\binom{9}{4} + \binom{9}{8} = 135$. By Theorem 24.6 we see that hyperbolic and elliptic quadratic forms have

$$\big(2^3+1\big)\frac{2^4-1}{2-1} = 135 \qquad \text{and} \qquad \big(2^4+1\big)\frac{2^3-1}{2-1} = 119$$

singular points respectively; so Q must be hyperbolic. Let $\mathcal{O}$ be the collection of all $9 = 2^3+1$ points spanned by vectors of weight 8, so that Q is hyperbolic. Let $\mathcal{O}$ be the set of all points $\langle x\rangle$ such that $wt(x) = 8$; then $wt(x+x') = 2$ for all $x \neq x'$ in $\mathcal{O}$, so that $\mathcal{O}$ is an ovoid.

It is known [62] that the $O^-_{2n}(q)$ quadric has no ovoids for $n \geq 3$; and [28] that the $O_{2n+1}(q)$ quadric has no ovoids for $n \geq 4$. Ovoids are known [19] to exist in $O^+_8(q)$ for all *prime* values of q, as well as for all $q \not\equiv 1 \mod 3$; see [33]. Ovoids of $O^+_8(q)$ (when they exist) are equivalent to spreads of $O^+_8(q)$, by a triality automorphism; see Section 24.9. It is *not* known whether the $O^+_8(25)$ quadric (the smallest open case) has any ovoids. The most significant open problem in this area is whether or not $O^+_{2n}(q)$ has any ovoids for $n \geq 5$. The most significant progress in this direction is the following, which for example proves the nonexistence of ovoids in $O^+_{10}(2^r)$ and $O^+_{10}(3^r)$. A **cap** in a polar space is a collection of points, no two collinear.

27.5 Theorem [6]**.** Every cap $\mathcal{O}$ in a nondegenerate quadric in $\mathbb{P}^n(\mathbb{F}_{p^r})$ has size

$$|\mathcal{O}| \leq \left[\tbinom{p+n-1}{n} - \tbinom{p+n-3}{n}\right]^r + 1.$$

In particular ovoids cannot exist unless

$$p^{\lfloor n/2\rfloor} > \tbinom{p+n-1}{n} - \tbinom{p+n-3}{n}.$$

Proof of Theorem 27.5. Actually we present here a proof of a slightly weaker bound. Let $k = |\mathcal{O}|$. For each $P \in \mathcal{O}$ the hyperplane $P^\perp \subset V$ contains P and no other point of $\mathcal{O}$; therefore the point-hyperplane incidence matrix A of $\mathbb{P}^n(\mathbb{F}_{p^r})$ has a $k \times k$ identity submatrix with rows indexed by points $P \in \mathcal{O}$ and columns indexed by hyperplanes $P^\perp$ for $P \in \mathcal{O}$. Since the p-rank of any matrix is bounded below by the p-rank of any of its submatrices, we have

$$|\mathcal{O}| = k \leq \operatorname{rank}_p A = \binom{p+n-1}{n}^r + 1$$

by Theorem 21.1. The slight improvement of this bound stated in the Theorem, arises from the observation that the $k \times k$ identity submatrix actually lies in a submatrix of A, having rows indexed by just the singular points. The rank of this submatrix is $\left[\binom{p+n-1}{n} - \binom{p+n-3}{n}\right]^r + 1$; see [6]. □

Exercises 27.

1. Let $\mathcal{O}$ be an ovoid in $O^+_{2n}(q)$, and let $\langle x\rangle$ be a singular point *not* in $\mathcal{O}$. Show that the hyperplane $x^\perp$ intersects $\mathcal{O}$ in exactly $q^{n-2}+1$ points, which give an ovoid in $x^\perp/\langle x\rangle \simeq O^+_{2n-2}(q)$. Thus conclude that if $O^+_{2n-2}(q)$ has no ovoids, then neither does $O^+_{2n}(q)$.

28. Generalized Quadrangles

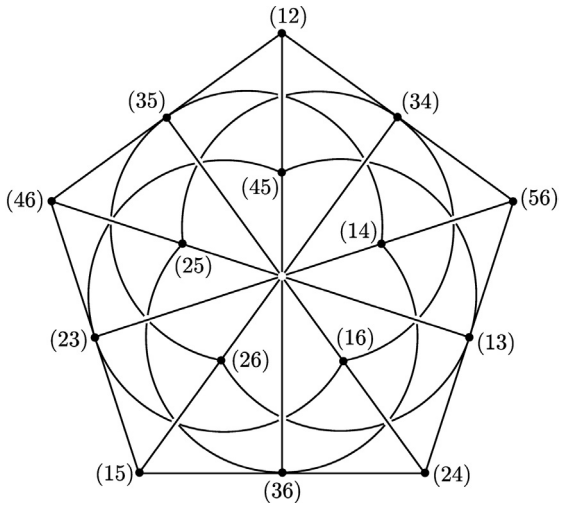
The symmetric group S_6 has 75 elements of order 2, which fall into three conjugacy classes:

(C1) (12), (13), ..., (56): $\binom{6}{2} = 15$ such elements;

(C2) (12)(34), (12)(35), ..., (34)(56): $3\binom{6}{4} = 45$ such elements;

(C3) (12)(34)(56), (12)(35)(46), ..., (16)(25)(34): $\frac{1}{6}\binom{6}{2}\binom{4}{2} = 15$ such elements.

Thus classes (C1) and (C3) consist of odd permutations; class (C2) consists of even permutations. Consider the incidence structure $(\mathfrak{P}, \mathfrak{L})$ with point and line sets given by the elements of order 2 in classes (C1) and (C3) respectively; and where a point $\sigma \in \mathfrak{P}$ lies on a line $\tau \in \mathfrak{L}$ precisely when σ and τ commute. One checks that this is a partial linear space, with three points on each line, and three lines through each point, as shown:



(15)(24)(36)	(13)(25)(46)	(16)(23)(45)	(12)(34)(56)	(14)(26)(35)	(14)(23)(56)	(12)(35)(46)	(13)(26)(45)	(16)(25)(34)	(15)(26)(34)	(16)(24)(35)	(12)(36)(45)	(14)(25)(36)	(13)(24)(56)	(15)(23)(46)	
		1	1			1					1			1	(12)
	1				1			1					1		(13)
1				1								1			(14)
1								1	1					1	(15)
	1						1			1				1	(16)
	1			1										1	(23)
1								1		1		1			(24)
	1				1			1				1			(25)
			1			1			1						(26)
			1		1			1		1					(34)
				1		1			1						(35)
1											1	1			(36)
		1					1				1			1	(45)
	1					1							1		(46)
			1		1							1			(56)

28.1 Figure. The Generalized Quadrangle of Order 2

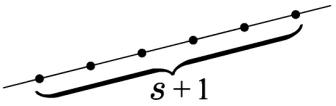
This is the unique generalized quadrangle of order 2. It is embedded in the projective plane of order 4, as follows: Let $\mathcal{O}$ be a hyperoval in $\mathbb{P}^2(\mathbb{F}_4)$, let $\mathfrak{P}$ be the $21 - 6 = 15$ points not in $\mathcal{O}$, and let $\mathfrak{L}$ be the $\binom{6}{2} = 15$ secants to $\mathcal{O}$. Then $(\mathfrak{P}, \mathfrak{L})$ is isomorphic to the structure defined above.

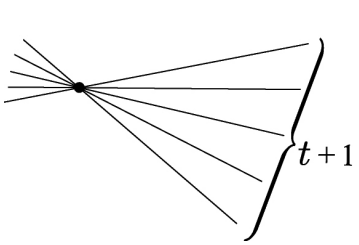
A **generalized quadrangle** is a point-line incidence structure satisfying the following:

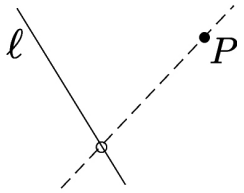
(GQ1) Every line has $s+1$ points, for some $s \geq 1$.

(GQ2) Every point lies on $t+1$ lines, for some $t \geq 1$.

(GQ3) If P is a point not on a line ℓ, then there is a unique line through P meeting ℓ.



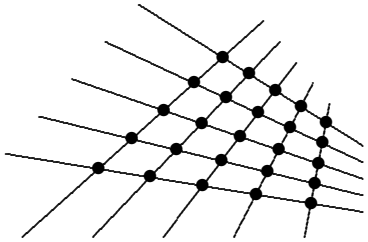




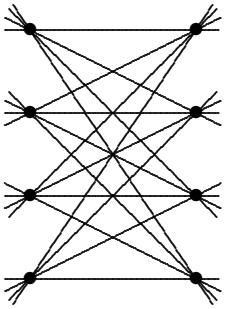
We call the pair (s, t) the **order** of the generalized quadrangle; or if $s = t$ we say simply a generalized quadrangle of order s. From the axioms it is clear that the point-line

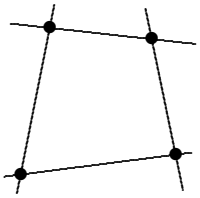
dual of a generalized quadrangle of order (s,t) is a generalized quadrangle of order (t,s). The example above is the unique generalized quadrangle of order 2, up to isomorphism; in particular since it is unique, it is self-dual (isomorphic to its dual). Usually one requires $s,t \geq 2$, giving the so-called **thick** generalized quadrangles; those with $s = 1$ or $t = 1$ are **thin**. Indeed a generalized quadrangle of order $(s,1)$ is just a 2-net of order $s+1$, i.e. a **grid** with $(s+1)^2$ points and $2(s+1)$ lines. A generalized quadrangle of order $(1,t)$ is a **dual grid** with two sets of $t+1$ points, and $(t+1)^2$ lines joining every point in one set with every pair in the other set. Some examples of thin generalized quadrangles are shown:

A thin generalized quadrangle (i.e. a grid) of order $(4,1)$



A thin generalized quadrangle (i.e. a dual grid) of order $(1,3)$





A thin generalized quadrangle of order 1, i.e. a quadrangle

Just as a 'projective plane of order 1' is a triangle, so a generalized quadrangle of order 1 is a quadrangle. So a projective plane may be reasonably viewed as a generalized triangle. In Section 29 we will describe generalized polygons from this perspective.

28.2 Proposition. A generalized quadrangle of order (s,t) has exactly $(s+1)(st+1)$ points and $(t+1)(st+1)$ lines.

Proof. Let P be a point. The set of all points is a disjoint union

$$\{P\} \cup \mathfrak{P}_1 \cup \mathfrak{P}_2$$

where $\mathfrak{P}_1$ is the set of all points distinct from P but collinear with P; and $\mathfrak{P}_2$ is the set of all remaining points. We have

$$|\mathfrak{P}_1| = (t+1)s$$

since each of the $(t+1)$ lines through P has s points of $\mathfrak{P}_1$, with no point counted twice. Every point $R \in \mathfrak{P}_2$ is collinear with exactly $t+1$ points in $\mathfrak{P}_1$, one on each of the lines through R, by (GQ3). Every point $Q \in \mathfrak{P}_1$ is collinear with st points in $\mathfrak{P}_2$: that's s points of $\mathfrak{P}_2$ on each of the t lines through Q other than the line PQ. So

$$|\mathfrak{P}_2| = \frac{|\mathfrak{P}_1|st}{t+1} = s^2t$$

and the total number of points is

$$1 + |\mathfrak{P}_1| + |\mathfrak{P}_2| \;=\; 1 + (t+1)s + s^2t \;=\; (s+1)(st+1).$$

The remaining conclusion follows by the dual argument. □

Just as there is an infinite family of classical projective planes coordinatized by fields, so also there are three infinite families of generalized quadrangles formed by classical polar spaces of rank 2 (i.e. having just two types of objects, points and lines) and their duals (so altogether six families). The more general polar spaces were defined in Section 24.10.

28.3 The $O_5(q)$ and $Sp_4(q)$ Quadrangles. We verify that the points and lines contained in an $O_5(q)$-quadric form a generalized quadrangle of order q. Certainly every line has $q+1$ points. If $\langle x\rangle$ is any singular point then by Theorem 24.3, $x^\perp/\langle x\rangle \simeq O_3(q)$ which is a projective plane having $q+1$ singular points (the points of a conic). These $q+1$ singular points correspond to $q+1$ totally singular lines in $x^\perp$ containing $\langle x\rangle$. This verifies (GQ1) and (GQ2).

Let $\langle x\rangle$ be a singular point, and let ℓ be a totally singular line not containing $\langle x\rangle$. Then $x^\perp$ is a hyperplane not containing ℓ, so this hyperplane $x^\perp$ meets ℓ in a single point $\langle y\rangle = x^\perp \cap \ell$. The totally singular line $\langle x, y\rangle$ is the unique line of the quadric containing $\langle x\rangle$ and meeting ℓ. Thus (GQ3) holds. This gives the $O_5(q)$ **quadrangle**, historically denoted $Q(4,q)$ since it is formed by the points and lines of a quadric in $\mathbb{P}^4(\mathbb{F}_q)$.

Recall that by Theorem 24.6, the $O_5(q)$-quadric has $(q^2+1)(q+1)$ points and the same number of lines, in agreement with Proposition 28.2. The $O_5(2)$ quadrangle is the unique generalized quadrangle of order 2 given as our first example above. Not surprisingly, the isometry group of the quadratic form of type $O_5(2)$ is isomorphic to S_6.

The $O_5(q)$ quadrangle is self-dual iff q is even. For arbitrary q, the dual of the $O_5(q)$ quadrangle is the **symplectic quadrangle** $Sp_4(q)$, historically denoted $W_3(q)$ or simply $W(q)$. This is simply the rank 2 symplectic polar space described in Section 24.10. An explicit duality between these two generalized quadrangles is given by the Klein correspondence κ. Recall that κ maps the lines of $\mathbb{P}^3(\mathbb{F}_q)$ to the points of the Klein quadric. It is not hard to check (using the exterior algebra by which κ was defined) that if we restrict κ to just the lines of an $Sp_4(q)$-quadrangle, the corresponding points of the quadric lie in an $O_5(q)$-hyperplane.

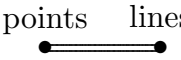
28.4 The $O_6^-(q)$ and $U_4(q)$ Quadrangles. Recall that the $O_6^-(q)$-quadric contains points and lines, but no planes. These points and lines form a quadrangle of order (q, q^2); the proof is similar to the previous case. This is the $O_6^-(q)$ **quadrangle**, also known as the $Q(5,q)$ **quadrangle**. Again by Theorem 24.6, the $O_6^-(q)$-quadric has $(q+1)(q^3+1)$ points and $(q^2+1)(q^3+1)$ lines, in agreement with Proposition 28.2.

The dual of the $O_6^-(q)$ quadrangle is a generalized quadrangle of order (q^2, q), which also has a classical construction. Here the points and lines are those lying in the Hermitian

surface $X_0^{q+1} + X_1^{q+1} + X_2^{q+1} + X_3^{q+1} = 0$ in $\mathbb{P}^3(\mathbb{F}_{q^2})$. This is the **unitary quadrangle** $U_4(q)$, also known as $H(3,q)$. It is simply the rank 2 unitary polar space of type $U_4(q)$ described in Section 24.10.

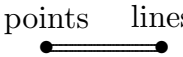
28.4 The $U_5(q)$ Quadrangles. The points and lines of $\mathbb{P}^4(\mathbb{F}_q)$ lying in the Hermitian variety $X_0^{q+1} + X_1^{q+1} + \cdots + X_4^{q+1} = 0$ form a generalized quadrangle of order (q^2, q^3), known as the **unitary quadrangle** $U_5(q)$; or historically, $H(4,q)$. This example also appeared in Section 24.10. Its dual is a generalized quadrangle of order (q^3, q^2), which however is not readily described other than as the dual of the quadrangle naturally arising from the Hermitian variety.

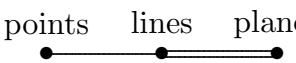
The Coxeter-Dynkin diagram for generalized quadrangles is the following diagram, whose symmetry (like that of the A_n diagram for projective spaces) indicates that the class of generalized quadrangles is closed under duality:

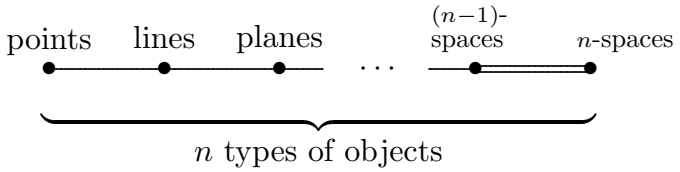
generalized quadrangle: 

(type B_2)

This diagram is the first member of the infinite sequence of diagrams, including the following:

B_2 polar space: 

B_3 polar space: 

B_n polar space: 

Polar spaces of type $O_{2n+1}(q)$, $O^-_{2n+2}(q)$, $Sp_{2n}(q)$, $U_{2n}(q)$ and $U_{2n+1}(q)$ are represented by the latter B_n diagram. Consider for example the points, lines and planes of the $O_7(q)$-quadric, with diagram B_3 shown above. Given a singular point $\langle x \rangle$, the totally singular lines and planes containing $\langle x \rangle$ form a geometry isomorphic to the totally singular lines and planes of $x^\perp/\langle x \rangle \simeq O_5(q)$ by Theorem 24.3. This agrees with the prediction of the Coxeter-Dynkin diagram, that the residue (Section 19) of the point $\langle x \rangle$ should be a geometry of type B_2 (the diagram formed by deleting the node for 'points' from the B_3 diagram). The residue of any plane is a geometry of type A_2, i.e. just a projective plane. And the residue of a line is a geometry of type $A_1 \oplus A_1$, i.e. a generalized digon.

This can be used for counting objects of each dimension in the associated polar space. For example, we know that the $O_7(q)$-quadric has $(q^6-1)/(q-1)$ points and $(q+1)(q^2+1)(q^3+1)$ planes, by Theorem 24.6. Let N be the number of lines in the quadric. Then the number of incident point-line pairs $(\langle x \rangle, \ell)$ in the quadric is $(q+1)N$. But for

each singular point $\langle x\rangle$, the quotient $x^\perp/\langle x\rangle \simeq O_5(q)$ is a generalized quadrangle of order q. There are $(q+1)(q^2+1)$ totally singular lines through $\langle x\rangle$, corresponding to the $(q+1)(q^2+1)$ points of the $O_5(q)$ quadrangle. Thus

$$(q+1)N = \frac{q^6-1}{q-1}(q+1)(q^2+1)\,,$$

which yields $N = (q^2+1)(q^6-1)/(q-1) = (q+1)(q^2+1)(q^4+q^2+1)$.

Next we give some constructions of nonclassical generalized quadrangles.

28.6 Example: Tits' Quadrangles. Let π_0 be a plane in $\mathbb{P}^3(\mathbb{F}_q)$, and let $\mathcal{O}$ be an oval in π_0. Consider the point-line incidence structure $(\mathfrak{P},\mathfrak{L})$ with $(q^2+1)(q+1)$ points given by

(i) the q^3 points of $\mathbb{P}^3(\mathbb{F}_q)$ outside π_0 (the 'affine' points),
(ii) the $q(q+1)$ planes $\pi \subset \mathbb{P}^3(\mathbb{F}_q)$ such that $\pi \cap \pi_0$ is a tangent line of $\mathcal{O}$, and
(iii) one additional point denoted ∞;

and $(q^2+1)(q+1)$ lines given by

(iv) the $q^2(q+1)$ lines ℓ of $\mathbb{P}^3(\mathbb{F}_q)$ such that $\ell \cap \pi_0$ is a point of $\mathcal{O}$, and
(v) the $q+1$ points of $\mathcal{O}$.

Incidence between points of type (i)–(iii) and lines of type (iv)–(v) is the natural containment, with the exception that the point ∞ of type (iii) is incident with all lines of type (v) and no lines of type (iv). It is straightforward to check that $(\mathfrak{P},\mathfrak{L})$ is a generalized quadrangle of order q. This quadrangle [25], discovered by Tits, is denoted $T_2(\mathcal{O})$. If $\mathcal{O}$ is a conic (as must be the case when q is odd, by Segre's Theorem), then $T_2(\mathcal{O})$ is isomorphic to the $O_5(q)$ quadrangle. Otherwise $T_2(\mathcal{O})$ is a new generalized quadrangle. The smallest nonclassical $T_2(\mathcal{O})$ arises for $q = 8$, where $\mathcal{O}$ is taken to be a 'pointed conic' (Section 12).

Similarly, we may take $S_0 \simeq \mathbb{P}^3(\mathbb{F}_q)$ to be a solid (i.e. hyperplane) of $\mathbb{P}^4(\mathbb{F}_q)$, and $\mathcal{O}$ an ovoid of the projective 3-space S_0. Consider the point-line incidence structure $(\mathfrak{P},\mathfrak{L})$ with $(q^3+1)(q+1)$ points given by

(i) the q^4 points of $\mathbb{P}^3(\mathbb{F}_q)$ outside S_0 (the 'affine' points),
(ii) the $q(q^2+1)$ solids $S' \subset \mathbb{P}^4(\mathbb{F}_q)$ such that $S \cap S_0$ is a tangent plane to $\mathcal{O}$ in S_0, and
(iii) one additional point denoted ∞;

and $(q^3+1)(q^2+1)$ lines given by

(iv) the $q^3(q^2+1)$ lines ℓ of $\mathbb{P}^4(\mathbb{F}_q)$ such that $\ell \cap S_0$ is a point of $\mathcal{O}$, and
(v) the q^2+1 points of $\mathcal{O}$.

Incidence between points of type (i)–(iii) and lines of type (iv)–(v) is the natural containment, with the exception that the point ∞ of type (iii) is incident with all lines of type (v)

and no lines of type (iv). Then $(\mathfrak{P}, \mathfrak{L})$ is a generalized quadrangle [25] of order (q, q^2), also discovered by Tits, and denoted $T_3(\mathcal{O})$. If $\mathcal{O}$ is an elliptic quadric (as must be the case when q is odd, by Barlotti's Theorem), then $T_3(\mathcal{O})$ is isomorphic to the $O_6^-(q)$ quadrangle. Otherwise $T_3(\mathcal{O})$ is a new generalized quadrangle. The smallest nonclassical $T_3(\mathcal{O})$ arises for $q = 8$, where $\mathcal{O}$ is taken to be a Suzuki-Tits ovoid.

28.7 Example: The Ahrens-Szekeres Quadrangles. The construction of Ahrens and Szekeres [1] gives nonclassical generalized quadrangles of order $(q-1, q+1)$ for any prime power q. Among the known generalized quadrangles, these parameters (s, t) (or their duals; see Payne's construction in the following example) in which s and t are not required to be powers of the same prime, or in fact prime powers at all. The construction takes different forms for q even and q odd. In both cases we denote the resulting generalized quadrangle by $AS(q)$ (although for most even values of q it is not uniquely determined by the parameter q).

For q even, the construction is a variation on Tits' $T_2(\mathcal{O})$ construction. Let π_0 be a plane in $\mathbb{P}^3(\mathbb{F}_q)$, and let $\mathcal{O}$ be a hyperoval in π_0 , so that $|\mathcal{O}|=q+2$. Consider the point-line incidence structure $(\mathfrak{P}, \mathfrak{L})$ whose q^3 points are all the points of $\mathbb{P}^3(\mathbb{F}_q)$ outside π_0 (i.e. the 'affine' points), and whose $q^2(q+2)$ lines are the lines ℓ of $\mathbb{P}^3(\mathbb{F}_q)$ such that $\ell \cap \pi_0$ is a point of $\mathcal{O}$. This gives $AS(q)$ for q even.

For q odd, fix a symplectic polarity $\perp$ of $\mathbb{P}^3(\mathbb{F}_q)$. Recall that the $(q^2+1)(q+1)$ points of this projective 3-space, and the $(q^2+1)(q+1)$ absolute lines, form an $Sp_4(q)$ quadrangle. Fix one point P of this quadrangle, and consider the incidence structure formed by the q^3 points of $\mathbb{P}^3(\mathbb{F}_q)$ not contained in the plane $P^\perp$, and the $q^2(q+2)$ lines given by

(i) the q^3+q^2 absolute lines ℓ (i.e. $\ell^\perp = \ell$) not containing P, and
(ii) the q^2 lines ℓ of $\mathbb{P}^3(\mathbb{F}_q)$ such that $\ell \cap P^\perp = \{P\}$.

This gives $AS(q)$ for q odd.

Note that $AS(2)$ is just a dual grid of order $(1, 3)$; and $AS(3)$ is isomorphic to $U_4(2)$, the unique generalized quadrangle of order $(2, 4)$. The smallest new quadrangle obtained by this construction is $AS(4)$, the unique generalized quadrangle of order $(3, 5)$.

28.8 Example: Payne's Quadrangles of Order $(q+1, q-1)$, q even. Let π_0 be a plane in $\mathbb{P}^3(\mathbb{F}_q)$ where q is even, and consider a hyperoval in π_0 partitioned as $\mathcal{O} \cup \{M, N\}$ where $\mathcal{O}$ is a q-arc. (It is known that for even $q > 4$, every q-arc extends to a unique hyperoval. So we must choose simply a hyperoval with two distinguished points M, N.) Consider the point-line incidence structure $(\mathfrak{P}, \mathfrak{L})$ whose $q^2(q+2)$ points are given by

(i) the q^3 points of $\mathbb{P}^3(\mathbb{F}_q)$ outside π_0 (i.e. the 'affine' points); and
(ii) the $2q^2$ planes of $\mathbb{P}^3(\mathbb{F}_q)$ containing exactly one of $\{M, N\}$;

and whose q^3 lines are the lines ℓ of $\mathbb{P}^3(\mathbb{F}_q)$ such that $\ell \cap \pi_0$ is a point of $\mathcal{O}$.

Payne [51] showed that this gives a generalized quadrangle of order $(q+1, q-1)$. In some, but not all cases, it is dual to $AS(q)$.

The following table lists the smallest known generalized quadrangles. For a more extensive table, together with further information on these examples, including explicit incidence matrices, see [47].

28.9 Table. Generalized Quadrangles of Small Order

order	no. of points	no. of lines	Name
$(2,2)$	15	15	$O_5(2),\ Sp_4(2)$
$(2,4)$	27	45	$U_4(2),\ AS(3)$
$(3,3)$	40	40	$O_5(3)$
$(3,3)$	40	40	$Sp_4(3)$
$(3,5)$	64	96	$AS(4)$
$(3,9)$	112	280	$U_4(3)$
$(4,2)$	45	27	$O_6^-(2)$
$(4,4)$	85	85	$O_5(4)$
$(4,6)$	125	175	$AS(5)$
$(4,8)$	165	297	$U_5(2)$

There are many open questions regarding existence and classification of generalized quadrangles with small parameters, for example: Does there exist a generalized quadrangle of order 6? or $(4,11)$? Is the generalized quadrangle of order $(4,6)$ unique?

A **spread** of a generalized quadrangle, is a collection of lines which partition the point set. Dually, an **ovoid** of a generalized quadrangle is a subset $\mathcal{O}$ of the points such that every line meets $\mathcal{O}$ in a unique point. In the notation of our first example, the generalized quadrangle of order 2 has $\{(12), (13), (14), (15), (16)\}$ as an ovoid and

$$\{(12)(36)(45),\ (13)(25)(46),\ (14)(23)(56),\ (15)(34)(26),\ (16)(24)(35)\}$$

as a spread.

28.10 Proposition. In a generalized quadrangle of order (s,t), every ovoid has $st+1$ points and every spread has $st+1$ lines.

Proof. Since a spread partitions the $(s+1)(st+1)$ points into lines each of size $s+1$, every spread must consist of

$$\frac{(s+1)(st+1)}{s+1} = st+1$$

lines. The dual argument gives the size of an ovoid. □

Here we summarize briefly what is known about existence and classification of ovoids and spreads in the classical generalized quadrangles:

- The $O_5(q)$ quadrangle always has *regular* ovoids arising from the embedding of the $O_4^-(q)$-quadric in $O_5(q)$. Other examples of ovoids are known only in characteristic 2 or 3. Spreads do not exist unless q is even, when the generalized quadrangle is self-dual (in which case spreads arise from spreads by duality).
- The $O_6^-(q)$ quadrangle has spreads but no ovoids.
- The $U_5(q)$ quadrangle has no ovoids. No spreads exist for $q = 2$, and for $q > 2$ the question of their existence is an open problem.

Of course each of the statements above can be dualized. For example the $Sp_4(q)$ quadrangle contains regular spreads for every q; indeed every regular spread of projective 3-space consists of totally isotropic lines with respect to some alternating bilinear form.

It is known that a thick generalized quadrangle with s and t both finite, must have $s \leq t^2$ and $t \leq s^2$. A particularly tantalizing open problem asks whether a generalized quadrangle can have $1 < s < \infty$ and t infinite. The answer is no for $s = 2$ (by a one-paragraph elementary proof [12, p.86]), for $s = 3$ (a nontrivial result of Brouwer [7]), and for $s = 4$ (a deep model-theoretic result of Cherlin [15]). For larger values of s the problem remains open.

Exercises 28.

1. Show that the collinearity graph of a generalized quadrangle is strongly regular, and find its parameters (v, r, λ, μ) in terms of the order (s, t) of the generalized quadrangle. (See Exercise#4.4 where the relevant terms from graph theory are defined.)

29. Generalized Polygons and Buildings

We give a graph-theoretic definition of generalized polygons, which requires the following terminology from graph theory. Let Γ be an ordinary graph (so Γ has no loops or multiple edges). A **path of length** n **from** v_0 **to** v_n in Γ is a sequence $(v_0, v_1, v_2, \ldots, v_n)$ of vertices in Γ such that v_{i-1} is adjacent to v_i for all $i = 1, 2, \ldots, n$; and where $v_0, v_1, \ldots, v_{n-1}$ are distinct. Such a path is **closed** if $n > 0$ and $v_0 = v_n$; evidently this requires $n \geq 3$. We say Γ is **connected** if for any two of its vertices, there exists a path from one vertex to the other. The **distance** from vertex v to vertex v' is the length of the shortest path from v to v' in Γ. This distance is finite whenever Γ is connected; otherwise it is undefined (or infinite). This defines a metric $d(v, v')$ on the vertices of Γ, i.e.

(D1) $d(v, v') \geq 0$, and equality holds iff $v = v'$;
(D2) $d(v', v) = d(v, v')$;
(D3) $d(v, v'') \leq d(v, v') + d(v', v'')$

for all vertices v, v', v''. The **diameter** of Γ is the maximum distance between any two of its vertices (or ∞ if Γ is not connected). The **girth** of Γ is the smallest n for which there exists a closed path $(v_0, v_1, \ldots, v_n)$ having $n \geq 3$ distinct vertices (so $v_0, v_1, \ldots, v_{n-1}$ are distinct and $v_n = v_0$). We agree that the girth of Γ is ∞, if no such closed path exists. We say Γ is **bipartite** if its vertices can be partitioned into two subsets V_1 and V_2, such that every edge of Γ has one endpoint in V_1 and the other endpoint in V_2. Note that Γ is bipartite iff every closed path in Γ has even length. Moreover if Γ is connected and bipartite, then the choice of partition $V_1 \cup V_2$ is unique, up to interchanging V_1 and V_2. Also Γ is **complete bipartite** if *every* vertex in V_1 is joined to every vertex in V_2 (in addition to no edges having both endpoints in V_i for the same i).

Recall that every point-line incidence structure $(\mathfrak{P}, \mathfrak{L})$ is equivalently described by its incidence graph Γ, the graph with vertex set $\mathfrak{P} \cup \mathfrak{L}$ and where each point-line pair (P, ℓ) is joined by an edge in Γ, iff the pair (P, ℓ) is incident in $(\mathfrak{P}, \mathfrak{L})$. It is understood here that $\mathfrak{P}$ and $\mathfrak{L}$ are disjoint sets; and that in Γ one clearly distinguishes which vertices correspond to points, and which vertices correspond to lines. Note that the dual incidence structure $(\mathfrak{L}, \mathfrak{P})$ yields the same incidence graph, but with the labelling of vertices (as points and lines) reversed. If Γ is connected then the only point-line incidence structures corresponding to Γ are $(\mathfrak{P}, \mathfrak{L})$ and $(\mathfrak{L}, \mathfrak{P})$.

A point-line incidence structure $(\mathfrak{P}, \mathfrak{L})$ is a **generalized n-gon** (where $n \geq 2$) if its incidence graph Γ satisfies the following:

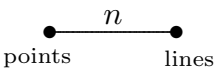
(GP1) The set of vertices of Γ is partitioned into 'points' and 'lines'. Every edge of Γ joins a point with a line; thus Γ is bipartite.

(GP2) Γ has girth n and diameter $2n$.

(GP3) Every line has $s + 1 \geq 2$ points, and every point is on $t + 1 \geq 2$ lines.

The **order** of the generalized polygon is the pair (s, t), or simply s if $s = t$. If $s = 1$ or $t = 1$ then the generalized polygon is **thin**. If $s, t \geq 2$ then the generalized polygon is **thick**.

The Coxeter-Dynkin diagram for a generalized n-gon is



(type $I_2(n)$)

which reduces to $A_1 \oplus A_1$, A_2, B_2 or G_2 when $n = 2, 3, 4, 6$ respectively. (The diagram is simplified by drawing an $(n-2)$-fold bond, rather than an n-fold bond.)

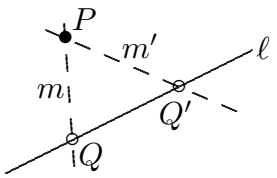
For every $n \geq 2$, an **n-gon** satisfies the conditions (GP1)–(GP3), giving the unique generalized n-gon of order 1, which is thin. This example has exactly n points and n lines, and Γ consists of simply a $2n$-cycle, i.e. a closed path of length $2n$. The thick examples are however more rare and much more interesting.

Consider first the case $n = 2$. Here we have a point set and a 'line' set (we should really call them blocks) such that every point is incident with every block; equivalently, Γ is a complete bipartite graph. To see this, note that Γ is bipartite by (GP1). If there

exists a point P and a line ℓ which are not incident, then since the distance from P to ℓ is odd, we must have $d(P,\ell) \geq 3$, contrary to the diameter of Γ being 2. Thus a generalized 2-gon of order (s,t) is the same thing as a generalized digon (Section 19) with $s+1$ points and $t+1$ lines.

Now consider the case $n = 3$. We claim that a generalized triangle (i.e. a generalized 3-gon) is simply a projective plane (or possibly a triangle, i.e. a thin projective plane, whose order is 1) and we must have $s = t$, the order of the plane. We now justify these conclusions. If two points P, Q are not joined by any line, then $d(P,Q) > 2$; and since $d(P,Q)$ is even, we have $d(P,Q) \geq 4$, contrary to the diameter of Γ being 3. If P and Q lie on two distinct lines ℓ and m then (P,ℓ,Q,m,P) is a closed path of length 4, violating the girth of Γ being 6. This proves the first projective plane axiom (P1), and (P2) follows dually. Note that the finite girth assumption rules out the closed configurations listed in Proposition 9.1(i)–(iv), leaving only cases (v) and (vi) of that Proposition. In fact (GP3) allows only the cases of a triangle or a projective plane.

Next consider the case $s = 4$. We claim that a generalized 4-gon is simply a generalized quadrangle as defined in Section 28. Let P be a point not on a line ℓ, so that $d(P,\ell) > 1$. Since $d(P,\ell)$ is odd and the diameter of Γ is 4, we must have $d(P,\ell) = 3$. This means there is a path (P,m,Q,ℓ) from P to ℓ in Γ. If there is *another* line m' joining P to ℓ, as shown, then we have a closed path (P,m,Q,ℓ,Q',m',P) of length 6, which is less than the girth 8, a contradiction. This verifies our claim.



It is interesting to note that thick generalized n-gons only exist for $n \in \{2,3,4,6,8\}$, and are apparently quite rare for $n \in \{6,8\}$.

29.1 Theorem (Feit and Higman [27]). Finite thick generalized n-gons exist only for $n \in \{2,3,4,6,8\}$.

The known generalized hexagons ($n = 6$) and generalized octagons ($n = 8$) are as follows:

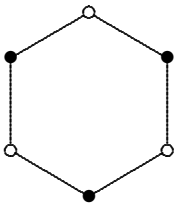
- Generalized hexagons of type $G_2(q)$ having order q for every prime power q, and their duals (these geometries are however self-dual for $q = 3^r$). Each such generalized hexagon has $(q^6-1)/(q-1)$ points and the same number of lines. The smallest example, $G_2(2)$, having 63 points and 63 lines, is the unique generalized hexagon of order 2.

- Generalized hexagons of type ${}^3D_4(q)$ having order (q, q^3) for every prime power q. These have $(q+1)(q^8+q^4+1)$ points and $(q^3+1)(q^8+q^4+1)$ lines. Also their duals. The smallest example, of order $(2, 8)$, has 819 points and 2457 lines.
- Generalized octagons of type ${}^2F_4(q)$ having order (q, q^2) for $q = 2^{2e+1}$. These have $(q+1)(q^3+1)(q^6+1)$ points and $(q^2+1)(q^3+1)(q^6+1)$ lines. Also their duals. The smallest example, of order $(2, 4)$, has 1755 points and 2925 lines.

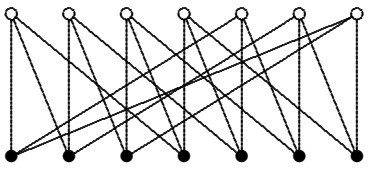
These families are named for the exceptional groups $G_2(q)$, ${}^3D_4(q)$, ${}^2F_4(q)$ of Lie type; see [63] for details. The smallest examples are explicitly provided in [47]. Despite much effort, no one has a clue why these should be the only generalized hexagons and octagons. In particular why should there not exist modifications of the families listed above, in the same way that the standard families for $n = 3, 4$ admit modifications. It is still an open question whether $G_2(3)$ is the unique generalized hexagon of order 3.

Even from a graph-theoretic viewpoint, generalized polygons (thought of in terms of their incidence graphs) are very interesting as they provide examples of sparse but highly connected graphs. The competing goals of being sparse but highly connected are desirable in the design of efficient communication networks, or for constructing good LDPC (low density parity check) codes. There is more than one way to formally express the condition 'sparse but highly connected'; and having large girth and small diameter does this quite nicely, although it is very hard to find explicitly constructible families of graphs which achieve these conditions.

Beginning in the 1950's, Jacques Tits defined a class of incidence structures known as *buildings* [57], [64], a notion which includes both projective spaces and polar spaces as special cases. A building of *rank* r has r types of objects. A building of rank 2 is simply a generalized polygon. In the general case, the structure of the building is described by a Coxeter-Dynkin diagram of the appropriate type, having r nodes, one node representing each type of object in the geometry. Edges of the graph specify the type of generalized polygons arising as rank 2 residues. A thin geometry corresponding to the given Coxeter-Dynkin diagram, is known as a *Coxeter complex.* A building (in the strict sense) is a thick geometry corresponding to the given diagram, which is a union of certain thin subgeometries (known as the *apartments* of the building) with the same diagram. Thus for example, any projective plane may be formed as a union of triangles:

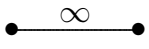


thin A_2 geometry (triangle)



thick A_2 geometry (projective plane)

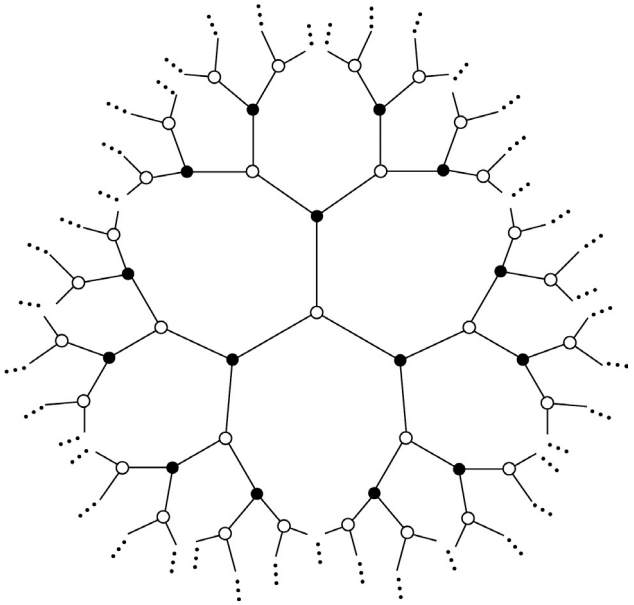
Another instructive example is the Coxeter diagram



(type $\widetilde{A}_1$)

whose buildings are infinite trees. Take for example a k-regular tree where $k \geq 2$: this is a (necessarily infinite) connected graph with no cycles (closed paths), in which every vertex has exactly k neighbours. This graph is bipartite; so we may choose an arbitrary vertex v to represent objects of type 1, and then any other vertex v' has type 1 or 2 according as the unique path from v to v' has even or odd length. Such a tree is the incidence graph of an $\widetilde{A}_1$-geometry. For $k = 2$ the geometry is thin, consisting of just an infinite path. For $k \geq 3$ the geometry is thick; it is in fact a building of type $\widetilde{A}_1$ whose apartments are infinite paths contained in the tree.

$k = 2$:

$k = 3$: 

Buildings were invented to study groups of Lie type. The groups of Lie type are classified by their Coxeter-Dynkin diagrams, and each such group is best understood by means of the natural geometry on which it acts: a building of the corresponding type. Thus for example, the group $PGL_{n+1}(F)$ corresponds to the Coxeter-Dynkin diagram of type A_n and has $\mathbb{P}^n(F)$ as its associated building.

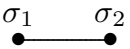
Consider a Coxeter-Dynkin diagram with r nodes labelled $1, 2, \ldots, n$, and edges labelled m_{ij}. Also set $m_{ii} = 1$. The group defined by the presentation

$$W = \langle s_1, s_2, \ldots, s_r \,:\, (s_i s_j)^{m_{ij}} = 1 \rangle$$

is the **Coxeter group** defined by the given diagram. It acts regularly on the objects of the associated Coxeter complex (i.e. thin building). In particular, the number of objects

in the Coxeter complex equals $|W|$. We illustrate with four Coxeter-Dynkin diagrams as examples:

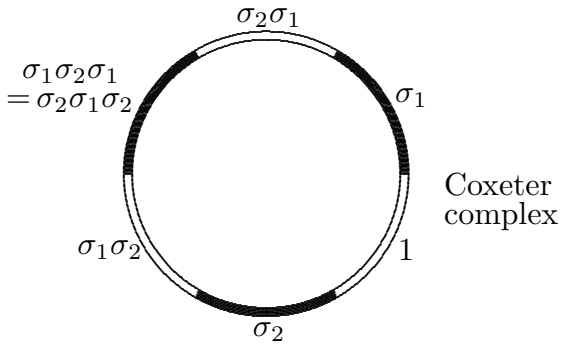
29.2 Example. Type A_2



$$(m_{ij}) = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$$

$W = \langle \sigma_1, \sigma_2 \,:\, \sigma_1^2 = \sigma_2^2 = (\sigma_1\sigma_2)^3 = 1 \rangle$
$= \{1,\ \sigma_1,\ \sigma_2,\ \sigma_1\sigma_2,\ \sigma_2\sigma_1,\ \sigma_1\sigma_2\sigma_1\}$

$W \cong S_3$ via $\sigma_1 \mapsto (12),\ \sigma_2 \mapsto (23)$

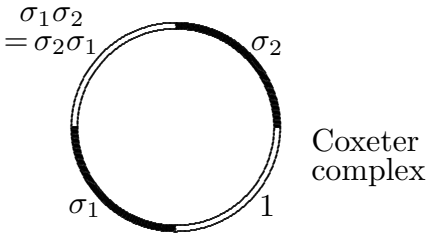


29.3 Example. Type $A_1 \oplus A_1$

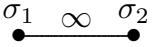
$$(m_{ij}) = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

$W = \langle \sigma_1, \sigma_2 \,:\, \sigma_1^2 = \sigma_2^2 = (\sigma_1\sigma_2)^2 = 1 \rangle$
$= \{1,\ \sigma_1,\ \sigma_2,\ \sigma_1\sigma_2\}$

W is elementary abelian

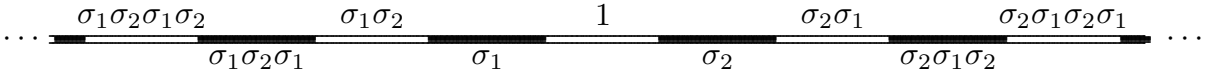


29.4 Example. Type $\widetilde{A}_1$

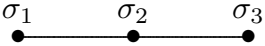


$W = \langle \sigma_1, \sigma_2 \,:\, \sigma_1^2 = \sigma_2^2 = 1 \rangle$ is infinite dihedral; $\sigma_1\sigma_2$ has infinite order

Coxeter complex:



29.5 Example. Type A_3

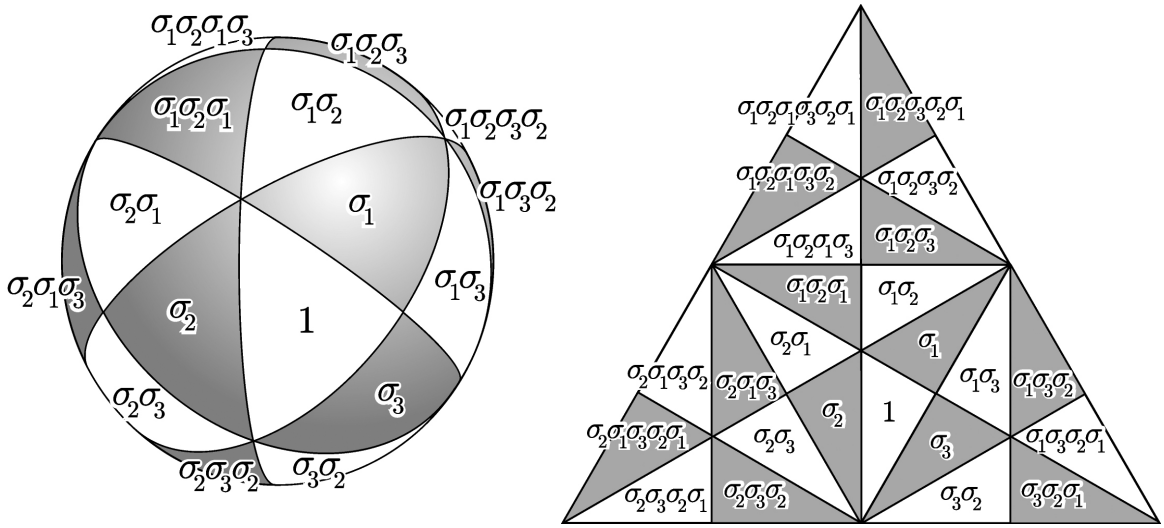


$$(m_{ij}) = \begin{bmatrix} 1 & 3 & 2 \\ 3 & 1 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

$W = \langle \sigma_1, \sigma_2, \sigma_3 \,:\, \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = (\sigma_1\sigma_2)^3 = (\sigma_1\sigma_3)^2 = (\sigma_2\sigma_3)^3 = 1 \rangle$

$W \cong S_4$ via $\sigma_1 \mapsto (12),\ \sigma_2 \mapsto (23),\ \sigma_3 \mapsto (34)$

The corresponding Coxeter complex is a triangulated 2-sphere:



This is in fact the geometry of the 3-simplex formed by all its subsimplices (Exercise #19.2). The 4 points, 6 lines, and 4 planes of this geometry are the subsets of $\{1, 2, 3, 4\}$ of size 1, 2, and 3 respectively. These give the 14 vertices of the triangulation. The 36 edges correspond to the incident pairs of objects (12 incident point-line pairs, 12 incident point-plane pairs, and 12 incident line-plane pairs). The 24 triangular faces correspond to the triples (P, ℓ, π) such that $P \in \ell \subset \pi$. We check that $14 - 36 + 24 = 2$, the Euler characteristic of the 2-sphere.

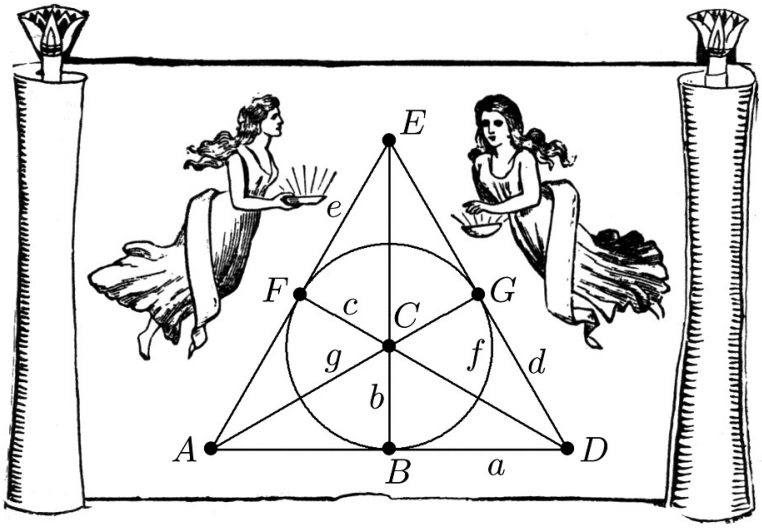
The Coxeter complex (or diagram, or matrix, or group), in turn, gives for each choice of field F a building, as a thickened-up version of the Coxeter complex. It also gives a presentation of the corresponding group G of Lie type defined over F, by means of generators and relations. In addition to the r generators of the Coxeter group (where r is the rank of the building), one generates G by means of certain subgroups whose elements are parameterized by F, and these elements must satisfy certain new relations as prescribed by the Coxeter group.

Coxeter gave a beautiful characterization of those diagrams for which W is finite. And in these cases, the associated Coxeter complex is a triangulated $(r-1)$-sphere, as the examples above suggest. Buildings corresponding to these diagrams are known as **spherical buildings**. Thus generalized n-gons are examples of spherical buildings. Tits, who first defined buildings axiomatically, also showed that with only certain exceptions of small rank, spherical buildings are classical: they are the buildings associated with groups of the corresponding Lie type. Of course the rank 2 buildings (generalized polygons) are far from classified, however. And even though higher rank buildings may in some sense be known, there remain many open problems regarding these geometries (for example, questions of existence of ovoids and spreads).

Although we've exhausted the time available this semester, we have really just arrived at the beginning of a very big subject that would require another course or two to introduce properly!

Exercises 29.

1. For a generalized hexagon of order (s, t), determine
   (a) the number of points and lines;
   (b) the size of an ovoid and the size of a spread.
   Justify your answers.
2. For a generalized octagon of order (s, t), determine
   (a) the number of points and lines;
   (b) the size of an ovoid and the size of a spread.
   Justify your answers.



Appendices

Appendix A1: Finite Fields

The ring of integers modulo a prime p is a field, denoted $\mathbb{F}_p$. For every $r \geq 1$ there exists a field of order $q = p^r$, which is unique up to isomorphism. This field, denoted $\mathbb{F}_q$, is an extension of $\mathbb{F}_p$ of degree r. It may be constructed as the quotient ring $\mathbb{F}_p[X]/(f(X))$ where $f(X) \in \mathbb{F}_p[X]$ is irreducible over $\mathbb{F}_p$ of degree r. Such an irreducible polynomial $f(X)$ of degree r exists; and although not unique, the quotient field $\mathbb{F}_p[X]/(f(X))$ is unique up to isomorphism.

The automorphism group of $\mathbb{F}_q$ is $\langle\sigma\rangle = \{1, \sigma, \sigma^2, \ldots, \sigma^{r-1}\}$, a cyclic group of order r where

$$\sigma : \mathbb{F}_q \to \mathbb{F}_q, \quad x \mapsto x^p.$$

The fixed field of σ is just $\mathbb{F}_p$, so that the extension $\mathbb{F}_q \supseteq \mathbb{F}_p$ is Galois; more generally the fixed field of σ^k is $\mathbb{F}_{p^d}$ where $d = gcd(k, r)$.

The multiplicative group $\mathbb{F}_q^\times = \mathbb{F}_q \setminus \{0\}$ of any finite field is cyclic: we have $\mathbb{F}_q^\times = \{1, \omega, \omega^2, \ldots, \omega^{q-2}\}$ for some $\omega \in \mathbb{F}_q^\times$ which is called a **primitive element**. In particular $\mathbb{F}_p[\omega] = \mathbb{F}_q$, which says that ω generates the field extension $\mathbb{F}_q \supseteq \mathbb{F}_p$ and so its minimal polynomial over $\mathbb{F}_p$, call it $f(X) = Irr_{\omega,\mathbb{F}_p}(X)$, is of degree r. But the condition that ω is a primitive element is stronger, and the resulting polynomial $f(X)$ is called a **primitive polynomial**.

A1.1 Example: The field of order 16. Let $f(X) = X^4 + X + 1 \in \mathbb{F}_2[X]$. This polynomial is irreducible. (If necessary this can be checked by considering all possible linear and quadratic factors, since altogether there are only six of these.) We denote by ω a zero of $f(X)$ in $\mathbb{F}_{16}$, so that

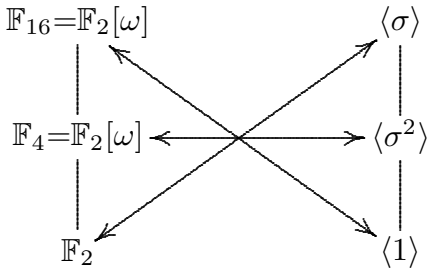
$$\begin{aligned}
\omega^4 &= \omega+1; & \omega^{10} &= \omega^4+\omega^2 &&= \omega^2+\omega+1;\\
\omega^5 &= \omega^2+\omega; & \omega^{11} &= \omega^3+\omega^2+\omega;\\
\omega^6 &= \omega^3+\omega^2; & \omega^{12} &= \omega^4+\omega^3+\omega^2 &&= \omega^3+\omega^2+\omega+1;\\
\omega^7 &= \omega^4+\omega^3 &&= \omega^3+\omega+1; & \omega^{13} &= \omega^4+\omega^3+\omega^2+\omega &&= \omega^3+\omega^2+1;\\
\omega^8 &= \omega^4+\omega^2+\omega &&= \omega^2+1; & \omega^{14} &= \omega^4+\omega^3+\omega &&= \omega^3+1;\\
\omega^9 &= \omega^3+\omega; & \omega^{15} &= \omega^4+\omega = 1.
\end{aligned}$$

Since $\mathbb{F}_{16} = \{0, 1, \omega, \omega^2, \omega^3, \ldots, \omega^{14}\}$, the element ω is primitive. Minimal polynomials for

each of the elements of $\mathbb{F}_{16}$ are listed as follows:

α	$Irr_{\alpha,\mathbb{F}_2}(X)$
$\omega, \omega^2, \omega^4, \omega^8$	X^4+X+1
$\omega^3, \omega^6, \omega^9, \omega^{12}$	$X^4+X^3+X^2+X+1$
ω^5, ω^{10}	X^2+X+1
$\omega^7, \omega^{11}, \omega^{13}, \omega^{14}$	X^4+X^3+1
1	$X+1$
0	X

Since $\mathbb{F}_{16}^{\times}$ is cyclic of order 15, it has $\phi(15) = 8$ generators; these are ω^k where $gcd(k, 15) = 1$. The minimal polynomials of these elements, namely X^4+X+1 and X^4+X^3+1, are the two primitive polynomials of degree 4 over $\mathbb{F}_2$. The elements $\omega^3, \omega^6, \omega^9, \omega^{12}$ are imprimitive but they are algebraic of degree 4, and so their minimal polynomial $X^4+X^3+X^2+X+1$ is also irreducible over $\mathbb{F}_2$; this makes altogether three irreducible polynomials of degree 4 over $\mathbb{F}_2$. The automorphism group of $\mathbb{F}_{16}$ is $\langle\sigma\rangle = \{1, \sigma, \sigma^2, \sigma^3\}$ and the fixed field of σ^2 is $\mathbb{F}_2[\omega^5] = \mathbb{F}_4$. The diagram of subfields, and the diagram of subgroups of $\langle\sigma\rangle$, are as shown:



The arrows show the Galois correspondence between each subgroup of $\mathrm{Aut}\,\mathbb{F}_{16}$ and its fixed subfield.

Consider an extension $E \supseteq F$ of finite fields, so that $E = \mathbb{F}_{q^n}$ and $F = \mathbb{F}_q$ for some prime power q, where $n = [E : F] \geq 1$. The group of all F-automorphisms of E is the Galois group

$$G = G(E/F) = \{1, \sigma, \sigma^2, \ldots, \sigma^{n-1}\}$$

where $\sigma : E \to E,\ x \mapsto x^\sigma$. (This generalizes the case $F = \mathbb{F}_p$ considered above.) The **norm** and **trace** maps of the extension $E \supseteq F$ are the maps

$$N_{E/F} : E \to F, \quad x \mapsto \prod_{g\in G} x^g = x^{1+\sigma+\sigma^2+\cdots+\sigma^{n-1}} = x^{1+q+q^2+\cdots+q^{n-1}}$$

and

$$T_{E/F} : E \to F, \quad x \mapsto \sum_{g\in G} x^g = x + x^\sigma + x^{\sigma^2} + \cdots + x^{\sigma^{n-1}} = x + x^q + x^{q^2} + \cdots + x^{q^{n-1}}.$$

The trace map $T_{E/F} : E \to F$ is F-linear and surjective; in particular

$$T_{E/F}(x+y) = T_{E/F}(x) + T_{E/F}(y).$$

The fact that $T_{E/F}$ is surjective follows from the fact that as a polynomial of degree $q^{n-1} < q^n$, the trace $T_{E/F}(x)$ cannot vanish at every element of E; and choosing $x \in E$ such that $T_{E/F}(x) \neq 0$, we see that $T_{E/F}(cx) = cT_{E/F}(x)$ takes on all values in F as we vary $c \in F$. The norm map $N_{E/F} : E \to F$ is multiplicative, i.e.

$$N_{E/F}(xy) = N_{E/F}(x)N_{E/F}(y).$$

It is also surjective, which we justify as follows. Since $N_{E/F}(0) = 0$, it suffices to consider the restriction of $N_{E/F}$ to the multiplicative group $E^\times$, a cyclic group of order $q^n - 1$. Since this map is $x \mapsto x^m$ where the exponent $m = 1+q+q^2+\cdots+q^{n-1} = (q^n-1)/(q-1)$ divides $|E^\times|$, the image is therefore the subgroup of order $q-1$, i.e. $F^\times$ as required.

A1.2 Matrix Representation of Fields. Let $E \supseteq F$ be a field extension of degree $n \geq 1$, and fix a basis $\{\omega_1, \ldots, \omega_n\}$ for E over F. For every $\alpha \in E$ the multiplication map $\mu_\alpha : E \to E$, $x \mapsto \alpha x$ is F-linear and so may be represented as an $n \times n$ matrix over F. Clearly $\mu_\alpha + \mu_\beta = \mu_{\alpha+\beta}$ and $\mu_\alpha\mu_\beta = \mu_{\alpha\beta}$ so the set of maps $\{\mu_\alpha : \alpha \in E\}$ (or just as well, the corresponding set of $n \times n$ matrices) is a ring, in fact a field, isomorphic to E via $\mu_\alpha \leftrightarrow \alpha$.

For example consider the extension $\mathbb{C} \supset \mathbb{R}$ with basis $\{1, i\}$; a typical element $\alpha = a+bi$ gives rise to the multiplication map

$$\mu_{a+bi}(x+yi) = (ax-by) + (bx+ay)i$$

which in matrix form appears as

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ -b & a \end{bmatrix} = \begin{bmatrix} ax-by & bx+ay \end{bmatrix}.$$

This gives a subring

$$R = \left\{ \begin{bmatrix} a & b \\ -b & a \end{bmatrix} : a, b \in \mathbb{R} \right\} \subset \mathbb{R}^{2\times 2}$$

where $\mathbb{R}^{2\times 2}$ denotes the ring of all 2×2 matrices over $\mathbb{R}$; moreover an isomorphism $\mathbb{C} \xrightarrow{\cong} R$ is given by $a+bi \mapsto \left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$.

More generally, consider any field extension $E \supseteq F$ of degree n and write $E = F[\omega] \cong F[X]/(f(X))$ where $f(X) \in F[X]$ is a monic polynomial of degree n which is irreducible over F, and $f(\omega) = 0$. We may write

$$f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0, \quad a_i \in F.$$

The F-linear transformation $\mu_\omega : E \to E$, $x \mapsto \omega x$ maps

$$\begin{aligned} 1 &\mapsto \omega; \\ \omega &\mapsto \omega^2; \\ \omega^2 &\mapsto \omega^3; \\ &\vdots \\ \omega^{n-2} &\mapsto \omega^{n-1}; \\ \omega^{n-1} &\mapsto \omega^n = -a_0 - a_1\omega - \cdots - a_{n-1}\omega^{n-1} \end{aligned}$$

and so the matrix of μ_ω relative to the basis $\{1, \omega, \omega^2, \ldots, \omega^{n-1}\}$ is given by

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & -a_3 & \cdots & -a_{n-1} \end{bmatrix}.$$

The matrix A is known as the **companion matrix** of $f(X)$. Since ω generates the extension $E \supseteq F$, the matrix A generates a subring $R \subseteq F^{n\times n}$ where $F^{n\times n}$ is the ring of all $n \times n$ matrices over F; moreover an isomorphism $E \xrightarrow{\cong} R$ is determined by $\omega \mapsto A$. With this identification of E with R, the norm and trace maps

$$N_{E/F} : E \to F, \qquad T_{E/F} : E \to F$$

of the extension $E \supseteq F$, defined as above, become simply the familiar determinant and trace maps

$$\det : R \to F, \qquad \mathrm{tr} : R \to F$$

of matrix theory.

Let us revisit the extension $\mathbb{F}_{16} \supset \mathbb{F}_2$ described in Example A1.1. In this case the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

generates a subring

$$R = \mathbb{F}_2[A] = \left\{ a{+}bA{+}cA^2{+}dA^3 \ : \ a, b, c, d \in \mathbb{F}_2 \right\} \subset \mathbb{F}_2^{4\times 4};$$

an isomorphism $\mathbb{F}_{16} \xrightarrow{\cong} R$ is given by

$$a{+}b\omega{+}c\omega^2{+}d\omega^3 \mapsto a{+}bA{+}cA^2{+}dA^3.$$

We see that finite extensions $E \supseteq F$ may be generally represented as rings of $n \times n$ matrices; and in principle this allows us to perform any desired computations in E explicitly. However in practice, the representation of E as a quotient ring $F[X]/(f(X))$ is much easier to implement, either by computer or by hand, since it represents elements of E as polynomials of degree $< n$, i.e. n-tuples over F; whereas $n \times n$ matrices require n^2 elements of F to represent every element of E. The matrix representation is primarily of conceptual value.

A1.3 Proposition. Consider a finite field $\mathbb{F}_q$. Recall that its nonzero elements $\mathbb{F}_q^\times$ form a cyclic multiplicative group of order $q-1$.

(i) Suppose q is odd. Then $\mathbb{F}_q^\times$ has $\frac{1}{2}\big(q-1\big)$ squares and the same number of non-squares. The element -1 is a square iff $q \equiv 1 \mod 4$. Every nonzero square has exactly two square roots in $\mathbb{F}_q$.

(ii) If q is even then every element of $\mathbb{F}_q$ has a unique square root in $\mathbb{F}_q$.

Proof. The set of nonzero squares is just the image of the group homomorphism $\phi : \mathbb{F}_q^\times \to \mathbb{F}_q^\times,\ x \mapsto x^2$. Note that $\mathbb{F}_q^\times$ is cyclic of order $q-1$, which is odd (or even) according as q is even (or odd). The result follows from elementary properties of cyclic groups. In particular for q odd, the unique element -1 of order two is a square, iff $\mathbb{F}_q^\times$ has an element of order 4, iff 4 divides $|\mathbb{F}_q^\times| = q-1$. □

Suppose $q = 2^r$ is even, so that $F = \mathbb{F}_q$ is an extension of $K = \mathbb{F}_2$. The map

$$F \to F, \quad x \mapsto x^2 + x$$

is K-linear. Denote the image of this map

$$\mathcal{C}_0 = \{x^2 + x \,:\, x \in \mathbb{F}_q\} \subset \mathbb{F}_q$$

which is a K-subspace of codimension 1. Also denote the complementary set

$$\mathcal{C}_1 = \mathbb{F}_q \smallsetminus \mathcal{C}_0 \,.$$

Note that $|\mathcal{C}_0| = |\mathcal{C}_1| = \frac{q}{2}$. Elements of $\mathcal{C}_0$ and $\mathcal{C}_1$ are called elements of **class 0** and **class 1** respectively.

A1.4 Theorem. Consider a quadratic polynomial $f(X) = aX^2 + bX + c \in \mathbb{F}_q[X]$ with $a \neq 0$.

(i) Suppose q is odd and let $\Delta = b^2 - 4ac$. Then $f(X)$ has two distinct linear factors, or one repeated linear factor, or is irreducible over $\mathbb{F}_q$, according as Δ is a nonzero square, or zero, or a nonsquare. If Δ is a square then the zeroes of $f(X)$ are

$$r_1, r_2 = \frac{-b \pm \sqrt{\Delta}}{2a} \in \mathbb{F}_q\,.$$

(ii) Suppose q is even. If $b = 0$ then $f(X) = a\big(X + \sqrt{\frac{c}{a}}\big)^2$ has a double zero $\sqrt{c/a}$. Now suppose $b \neq 0$. Then $f(X)$ has two distinct zeroes, or is irreducible over $\mathbb{F}_q$, according as $\frac{ac}{b^2}$ is of class 0 or 1 respectively. If $r \in \mathbb{F}_q$ is a zero then the other zero is $r + \frac{b}{a}$.

Proof. The case of odd characteristic is well-known and the proof is the same as in characteristic zero. So consider $q = 2^e$. The case $b = 0$ is clear, so assume $b \neq 0$. The condition for an element $r \in \mathbb{F}_q$ to be a zero of $f(X)$ is that

$$u^2 + u = \frac{ac}{b^2}, \qquad \text{where } u = \frac{ar}{b}\,.$$

Consider the trace map $T = T_{\mathbb{F}_q/\mathbb{F}_2} : \mathbb{F}_q \to \mathbb{F}_2,\ t \mapsto t + t^2 + t^4 + \cdots + t^{2^{e-1}}$. Note that

$$T(u^2) = u^2 + u^4 + \cdots + x^{2^{e-1}} + x = T(u)$$

since $t^q = t$ for all $t \in \mathbb{F}_q$. So a necessary condition for $f(X)$ to have zeroes in $\mathbb{F}_q$ is that

$$T\big(\tfrac{ac}{b^2}\big) = T(u^2) + T(u) = 2T(u) = 0,$$

i.e. $\frac{ac}{b^2}$ must be of class 0. If r is indeed a zero of $f(X)$ then clearly

$$f(X) \;=\; aX^2 + bX + c \;=\; a\big(X + r\big)\big(X + r + \tfrac{b}{a}\big).$$

□

By convention we agree that $0^0 = 1$.

A1.5 Lemma. Let $k \in \{0, 1, 2, \ldots, q-1\}$. Then

$$\sum_{a \in \mathbb{F}_q} a^k = \begin{cases} 0, & \text{if } k < q-1; \\ -1, & \text{if } k = q-1. \end{cases}$$

Proof. For $k = q-1$ all terms in the sum are 1 except the term $0^{q-1} = 0$, so the sum is $q-1 = -1$ (in $\mathbb{F}_q$). Now suppose $k \in \{0, 1, 2, \ldots, q-2\}$ and denote $S_k = \sum_{a\in\mathbb{F}_q} a^k$. For every $\lambda \in \mathbb{F}_q^\times$ the map $\mathbb{F} \to \mathbb{F}$, $x \mapsto \lambda x$ is a permutation so

$$S_k = \sum_{a\in\mathbb{F}_q} a^k = \sum_{a\in\mathbb{F}_q} (\lambda a)^k = \lambda^k \sum_{a\in\mathbb{F}_q} a^k = \lambda^k S_k \,.$$

Now $(\lambda^k - 1)S_k = 0$ for all $\lambda \in \mathbb{F}_q^\times$. Since the polynomial $X^k - 1$ has at most $k \le q-2$ zeroes there exists $\lambda \in \mathbb{F}_q^\times$ such that $\lambda^k \ne 1$, and this forces $S_k = 0$. □

A1.6 Theorem (Chevalley-Warning). Consider a system of polynomials $f_i(X_1, \ldots, X_n) \in \mathbb{F}_q[X_1, X_2, \ldots, X_n]$ for $i = 1, 2, \ldots, k$ of total degree $\sum_i \deg f_i(X_1, \ldots, X_n) < n$. Let $\mathcal{S}$ be the set of all simultaneous zeroes of the f_i's in $\mathbb{F}_q^n$, i.e.

$$\mathcal{S} = \big\{(a_1, a_2, \ldots, a_n) \in \mathbb{F}_q^n \,:\, f_i(a_1, a_2, \ldots, a_n) = 0 \text{ for all } i = 1, 2, \ldots, k\big\}.$$

Then $|\mathcal{S}|$ is divisible by p, where $q = p^r$.

Proof. We first extend Lemma A1.5 by observing that if $k_1, k_2, \ldots, k_n \ge 0$ with $\sum_i k_i < (q-1)n$, then

$$\sum_{a_1,a_2,\ldots,a_n\in\mathbb{F}_q^n} a_1^{k_1} a_2^{k_2} \cdots a_n^{k_n} \;=\; \prod_{1\le i\le n} \sum_{a\in\mathbb{F}_q} a^{k_i} = 0;$$

this holds since at least one of the exponents $k_i < q-1$. Now consider the polynomial

$$f(X_1, X_2, \ldots, X_n) \;=\; \prod_{1\le i\le k} \big(1 - f_i(X_1, X_2, \ldots, X_n)^{q-1}\big) \;\in \mathbb{F}_q[X_1, X_2, \ldots, X_n]$$

of total degree $(q-1)\sum_i k_i < (q-1)n$. Clearly

$$f(a_1, a_2, \ldots, a_n) = \begin{cases} 1, & \text{if } (a_1, a_2, \ldots, a_n) \in \mathcal{S}; \\ 0, & \text{otherwise.} \end{cases}$$

By considering each monomial term in the multinomial expansion of $f(X_1, X_2, \ldots, X_n)$, we have

$$\sum_{a_1,a_2,\ldots,a_n\in\mathbb{F}_q^n} f(a_1, a_2, \ldots, a_n) = 0 \in \mathbb{F}_q \,.$$

Since the sum has $|\mathcal{S}|$ ones and all other terms are zero, the result follows. □

Let $f : \mathbb{F}_q \to \mathbb{F}_q$. Then f can be uniquely expressed as a polynomial in $\mathbb{F}_q[X]$ of degree less than q, for example using Lagrange interpolation. A polynomial $f(X) \in$

$\mathbb{F}_q[X]$ of degree less than q is called a **permutation polynomial** if the function $\mathbb{F}_q \to \mathbb{F}_q$ represented by f is bijective. We give Dickson's condition for a polynomial to be a permutation polynomial.

First observe that if $f(X) \in \mathbb{F}_q[X]$ is any polynomial, we may divide $f(X)$ by $X^q - X$ to obtain $f(X) = (X^q - X)g(X) + r(X)$ where $\deg r(X) < q$. Then $r(X)$ is the unique polynomial of degree less than q which represents the same function $\mathbb{F}_q \to \mathbb{F}_q$ as the original $f(X)$. We call $r(X)$ the **reduction** of $f(X)$ mod $(X^q - X)$.

A1.7 Theorem (Dickson's Criterion). Let $f(X) \in \mathbb{F}_q[X]$ be a polynomial of degree less than q. Then $f(X)$ is a permutation polynomial iff the following two conditions hold:

(i) $f(X)$ has exactly one zero in F; and

(ii) for every $k \in \{1, 2, \ldots, q-2\}$, the reduction of $f(X)^k$ mod $(X^q - X)$ has degree at most $q-2$.

Proof. First suppose $f(X)$ is a permutation polynomial, so that (i) holds. For each $k \in \{0, 1, 2, \ldots, q-2\}$ write

$$f(X)^k \equiv \sum_{0 \le i < q} b_{ki} X^i \mod (X^q - X)$$

where $b_{ki} \in \mathbb{F}_q$. By Lemma A1.5 we have

$$0 = \sum_{a \in F} f(a)^k = \sum_{0 \le i < q} b_{ki} \sum_{a \in \mathbb{F}_q} a^i = b_{k,q-1}$$

for $k = 0, 1, 2, \ldots, q-2$.

Conversely, suppose (i) and (ii) hold. For each $a \in \mathbb{F}_q$ let $n_a = |f^{-1}(a)|$, which is the number of solutions of $f(x) = a$ for $x \in \mathbb{F}_q$. But

$$\text{(A1.8)} \qquad \sum_{a \in \mathbb{F}_q} n_a a^k = \left\{ \begin{array}{ll} 0, & \text{for } k = 0, 1, 2, \ldots, q-2; \\ -1, & \text{for } k = q-1 \end{array} \right\} = \sum_{a \in \mathbb{F}_q} a^k.$$

The validity of (A1.8) follows from (i) for $k = q-1$, and from (ii) for $k = 0, 1, 2, \ldots, q-2$ by Lemma A1.5. For each $b \in \mathbb{F}_q$ consider the polynomial

$$g_b(X) = 1 - (X - b)^{q-1} = \sum_{0 \le k < q} c_{bk} X^k \in \mathbb{F}_q[X].$$

For each $a \in \mathbb{F}_q$ we see that

$$g_b(a) = \delta_{ab} = \begin{cases} 1, & \text{if } a = b; \\ 0, & \text{otherwise.} \end{cases}$$

Clearly

$$\sum_{0\le k<q} c_{bk} \sum_{a\in\mathbb{F}_q} (n_a - 1)a^k = 0$$

since the inner sum vanishes by (A1.8); but after interchanging the order of summation we obtain

$$0 = \sum_{a\in\mathbb{F}_q} (n_a - 1) \sum_{0\le k<q} c_{bk} a^k = \sum_{a\in\mathbb{F}_q} (n_a - 1)\delta_{ab} = n_b - 1 \in \mathbb{F}_q$$

which says that $n_b \equiv 1 \bmod p$ for all $b \in \mathbb{F}_q$. Since the n_b's are non-negative integers whose sum is q, this clearly implies that $n_b = 1 \in \mathbb{Z}$ for all $b \in \mathbb{F}_q$. □

Appendix A2: Groups

A2.1 Permutation Groups

Let X be a set. The collection of all bijections $X \to X$ is a group under composition, called the **symmetric group** on X, denoted $\operatorname{Sym} X$. Elements of $\operatorname{Sym} X$ are called **permutations** of X. Permutations are often written using standard cycle notation, especially when $X = \{1, 2, \ldots, n\}$, in which case we abbreviate $S_n = \operatorname{Sym}\{1, 2, \ldots, n\}$. The action of a permutation $\sigma \in \operatorname{Sym} X$ on X is denoted either using superscripts, as in $x \mapsto x^\sigma$ (**right action**); or using the usual function notation $x \mapsto \sigma(x)$ (**left action**), depending on context. We prefer right action, unless this results in too many nested subscripts and superscripts. Multiplication in $\operatorname{Sym} X$ is left-to-right composition in the case of right actions, so that $(x^\sigma)^\tau = x^{\sigma\tau}$; or right-to-left composition in the case of left actions, so that $(\sigma \circ \tau)(x) = \sigma(\tau(x))$.

A **permutation group** on X is a subgroup $G \leq X$. More generally given an arbitrary group G, a **permutation action** or **permutation representation** of G on X is a homomorphism $\theta : G \to \operatorname{Sym} X$. Such an action is **faithful** if θ is injective, in which case $G \cong \theta(G)$ so that we may identify G with the permutation group $\theta(G) \leq \operatorname{Sym} X$. If G acts on X via θ, we denote the action of an individual element $g \in G$ by $x \mapsto x^{\theta(g)}$ or simply $x \mapsto x^g$ if the choice of action θ is clear from context. The **degree** of G is $|X|$, the number of points being permuted.

Let G be a permutation group on a set X, or more generally a group acting on a set X. For each $x \in X$, the **stabilizer** of x is the subgroup

$$G_x = \{g \in G \,:\, x^g = x\}.$$

The **orbit** of x is the subset

$$x^G = \{x^g \,:\, g \in G\} \subseteq X.$$

It is well-known that the orbits of G form a partition of X; and that for every $x \in X$ the index of the stabilizer equals the size of the orbit:

A2.2 Theorem. If G acts on X and $x \in X$, then $[G : G_x] = |x^G|$.

Assume G is a *finite* group acting on a *finite* set X, and let $x_1, \ldots, x_n \in X$ be representatives of the distinct orbits of G on X. The set of points of X fixed by $g \in G$ is denoted by

$$\operatorname{Fix}_X(g) = \{x \in X \,:\, x^g = x\}.$$

A2.3 Theorem. The average number of points of X fixed by elements of G, equals n, the number of orbits. That is,

$$\tfrac{1}{|G|}\sum_{g\in G}|\mathrm{Fix}_X(g)| = n.$$

Proof. Let $\mathcal{S}$ be the set of pairs (x,g) with $x\in X,\ g\in G$ and $x^g = x$. We count in two different ways the number of such pairs. On the one hand

$$|\mathcal{S}| = \sum_{g\in G}|\mathrm{Fix}_X(g)|.$$

On the other hand, for each $x\in x_i^G$ the number of elements of G fixing x is

$$|G_x| \;=\; \frac{|G|}{|x^G|} \;=\; \frac{|G|}{|x_i^G|}\,;$$

thus

$$\begin{aligned}
|\mathcal{S}| &= \sum_{x\in X}|G_x| \\
&= \sum_{1\le i\le n}\ \sum_{x\in x_i^G}|G_x| \\
&= \sum_{1\le i\le n}\ \sum_{x\in x_i^G}\frac{|G|}{|x_i^G|} \\
&= \sum_{1\le i\le n}|x_i^G|\cdot\frac{|G|}{|x_i^G|} \\
&= \sum_{1\le i\le n}|G| \\
&= n|G|.
\end{aligned}$$

□

A permutation group G (or more generally, an action of a group G) on X is **transitive** if there is only one orbit; that is, if for all $x,x'\in X$ there exists $g\in G$ such that $x^g = x'$. We also say G is **sharply transitive**, or **regular**, if the choice of $g\in G$ mapping $x\mapsto x'$ is unique. Note that G permutes X regularly iff $|G| = |X|$ and every point $x\in X$ has trivial stabilizer $G_x = 1$. We say that G is **doubly transitive** or **2-transitive** on X if G is transitive on ordered pairs in X, i.e. if for all pairs (x,y) and (x',y') with $x\neq y$ and $x'\neq y'$ in X, there exists $g\in G$ such that $(x^g,y^g) = (x',y')$. If moreover the choice of $g\in G$ is unique, we say G is **sharply 2-transitive** on X.

The following is often useful in determining the size and structure of an automorphism group.

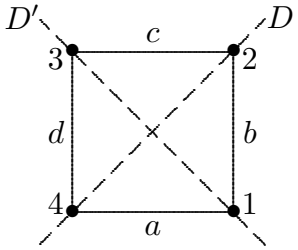
A2.4 Theorem. Let G be a group acting on a set X, and suppose $H \leq G$ is a transitive subgroup. Then for every element $x \in X$, the stabilizer G_x satisfies

$$G = G_x H = \{kh : k \in G_x,\ h \in H\}.$$

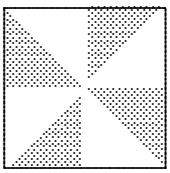
Proof. Let $g \in G$. Since H is transitive, there exists $h \in H$ mapping $x \mapsto x^g$. Then h^{-1} maps $x^g \mapsto x$, i.e. $x^{gh^{-1}} = x$. Thus $gh^{-1} \in G_x$ and $g = (gh^{-1})h \in G_x H$. □

Let $\theta : G \to \operatorname{Sym} X$ and $\psi : G \to \operatorname{Sym} Y$ be two permutation representations of the same group G, on possibly different sets X, Y. We say these two permutation representations are **equivalent** if there exists a bijection $f : X \to Y$ such that $f(x^{\theta(g)}) = f(x)^{\psi(g)}$ for all $x \in X$ and $g \in G$. Unravelling this definition, we see that two actions of G are equivalent iff one is obtained from the other by simply renaming the points that are being permuted. This is the natural equivalence relation for permutation groups and actions, just as isomorphism is the natural equivalence relation for abstract groups.

For example, let G be the symmetry group of the square in the Euclidean plane as shown, so that G is dihedral of order 8:



Then G has several natural actions, including an action θ on the set of vertices $X = \{1, 2, 3, 4\}$; and an action ψ on the set of four sides $Y = \{a, b, c, d\}$; and an action π on the set of diagonals $Z = \{D, D'\}$. No two of these actions are equivalent. For example, let g be the reflection in the diagonal D; then the action of g on vertices is given by $\theta(g) = (13)$, whereas the action on edges is given by $\psi(g) = (ab)(cd)$. Since $\theta(g)$ is a transposition whereas $\psi(g)$ is a product of two disjoint transpositions, the two actions are inequivalent. The third action π is not even faithful; four of the elements of G act trivially on Z (i.e. they both diagonals) whereas the other four act as $(D\,D')$. Note that the two diagonals and the other two axes of symmetry divide the square into eight isosceles triangles, as shown:



Then G regularly permutes these eight triangles.

Every group G has a regular permutation action on the elements of $X = G$ by right-multiplication; this is Cayley's Representation Theorem. More generally, let $H \leq G$ be any subgroup, and consider the set $H\backslash G = \{Hg : g \in G\}$ of right cosets of H in G. Then G acts transitively on $X = H\backslash G$ by right-multiplication; moreover every transitive permutation action of G is equivalent to such an action for some choice of subgroup $H \leq G$. Indeed, given a transitive permutation action of G on X, let $x \in X$ and $H = G_x$; then the action of G on X is equivalent to the action of G on $G_x\backslash G$. In the case $G_x = 1$, we recover the regular representation.

A2.5 Linear Groups

Let F be a field. The **general linear group** $GL_n(F)$ is the multiplicative group consisting of all invertible $n \times n$ matrices over F. Its centre is the subgroup of all **scalar matrices**, i.e. nonzero scalar multiples of the identity matrix; thus $Z(GL_n(F))$ is naturally isomorphic to $F^\times$, the multiplicative group of all nonzero scalars. The **projective general linear group** is the quotient group

$$PGL_n(F) = GL_n(F)/Z(GL_n(F)).$$

Its elements may be thought of as invertible matrices, but with two such matrices identified whenever one is a nonzero scalar multiple of the other. The group $PGL_n(F)$ has a natural permutation action on the set of all 1-dimensional subspaces of F^n, since multiplication by a nonzero scalar fixes every 1-dimensional subspace. Similarly for each $k \in \{1, 2, \ldots, n-1\}$, the group $PGL_n(F)$ has a natural permutation action on the set of all k-dimensional subspaces of F^n. We easily check that this action is faithful.

Now suppose $F = \mathbb{F}_q$. The number of invertible $n \times n$ matrices over $\mathbb{F}_q$ is

$$|GL_n(\mathbb{F}_q)| = (q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{n-1})$$

since if $A = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ is such a matrix, where $v_1, \ldots, v_n$ are the columns of A, then there are

$q^n - 1$ choices of v_1, i.e. any nonzero vector of length n;
$q^n - q$ choices of v_2 for each v_1, i.e. any vector not a multiple of v_1;
$q^n - q^2$ choices of v_3 for each (v_1, v_2), i.e. any vector not in $\langle v_1, v_2 \rangle$;

etc.; and finally

$q^n - q^{n-1}$ choices of v_n for each $(v_1, v_2, \ldots, v_{n-1})$, i.e. any vector not in $\langle v_1, v_2, \ldots, v_{n-1} \rangle$.

Consequently

$$|PGL_n(\mathbb{F}_q)| = \tfrac{1}{q-1}(q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{n-1}).$$

For example we see that $|PGL_2(\mathbb{F}_2)| = 6$. Since this group faithfully permutes the three 1-dimensional subspaces of $\mathbb{F}_2^2$ (the $\begin{bmatrix}2\\1\end{bmatrix}_2 = 3$ points of $\mathbb{P}^1(\mathbb{F}_2)$) we must have $PGL_2(\mathbb{F}_2) \cong S_3$. In fact $GL_2(\mathbb{F}_2) \cong S_3$ since the centre of $GL_2(\mathbb{F}_2)$ is trivial.

In the same way, $|PGL_2(\mathbb{F}_3)| = 24$. From the action of $PGL_2(\mathbb{F}_3)$ on the four points of $\mathbb{P}^2(\mathbb{F}_3)$ we see that $PGL_2(|F_3) \cong S_4$.

Similarly, $|PGL_2(\mathbb{F}_4)| = 60$ and $PGL_2(\mathbb{F}_4)$ permutes the 5 points of $\mathbb{P}^2(\mathbb{F}_4)$. Since S_5 has a unique subgroup of order 60, we obtain $PGL_2(\mathbb{F}_4) \cong A_5$, the alternating group of degree 5.

Let V be a vector space over a field F. A map $T : V \to V$ is **semilinear** if there exists an automorphism $\sigma \in \operatorname{Aut} F$ such that

$$T(au + bv) = a^\sigma T(u) + b^\sigma T(v) \quad \text{for all } u, v \in V;\ a, b \in F.$$

The case $\sigma = 1$ gives simply a linear transformation as a special case. If $\sigma \in \operatorname{Aut}\mathbb{C}$ is complex conjugation, we call T simply a **conjugate linear transformation**. The group of all semilinear transformations $V \to V$ is denoted $\Gamma L(V)$, or $\Gamma L_n(F)$ if $V = F^n$. Again we write

$$P\Gamma L(V) = GL(V)/Z$$

where $Z \leq P\Gamma L(V)$ is the normal (although not central) subgroup consisting of all scalar maps $v \mapsto \lambda v$ where $0 \neq \lambda \in F$. We have

$$|\Gamma L_n(\mathbb{F}_q)| = r|GL_n(\mathbb{F}_q)|; \qquad |P\Gamma L_n(\mathbb{F}_q) = r|PGL_n(\mathbb{F}_q)|$$

where $q = p^r$ and the orders of $GL_n(\mathbb{F}_q)$ and $PGL_n(\mathbb{F}_q)$ are listed above. Every $T \in \Gamma L(V)$ is uniquely representable in the form

$$v \mapsto Av^\sigma$$

where $A \in GL(V)$ and $\sigma \in \operatorname{Aut} F$ acts coordinatewise on V; thus $(v_1, v_2, \ldots, v_n)^\sigma = (v_1^\sigma, v_2^\sigma, \ldots, v_n^\sigma)$.

An **affine linear transformation** $T : V \to V$ is a map of the form

$$T_{A,a}(v) = Av + a$$

for some $A \in GL(V)$ and $a \in V$. The group of all affine linear transformations $V \to V$ is the **affine general linear group** denoted $AGL(V)$, or $AGL_n(F)$ if $V = F^n$. The right-to-left composition law on $AGL(V)$ is

$$T_{A,a} \circ T_{B,b} = T_{AB,\, Ab+a}\,.$$

It follows that $AGL_n(F)$ is isomorphic to the subgroup

$$\left\{ \begin{bmatrix} A & a^T \\ 0 & 1 \end{bmatrix} \,:\, A \in GL_n(F),\ a \in F^n \right\} < GL_{n+1}(F)$$

and the map $T_{A,a} \mapsto \left[\begin{smallmatrix} A & a^T \\ 0 & 1 \end{smallmatrix}\right]$ is an isomorphism. (Here we think of F^n as column vectors of length n.) We can go one step further and define $A\Gamma L(V)$ as the group of **affine semilinear transformations**

$$v \mapsto Av^\sigma + a$$

where $A \in GL(V)$, $\sigma \in \operatorname{Aut} F$ and $a \in V$. Note that

$$|AGL_n(\mathbb{F}_q)| = q^n |GL_n(\mathbb{F}_q)|; \qquad |A\Gamma L_n(\mathbb{F}_q)| = rq^n |GL_n(\mathbb{F}_q)|$$

with notation as above.

One often abbreviates $GL_n(q) = GL_n(\mathbb{F}_q)$ and similarly for the other matrix groups over $\mathbb{F}_q$.

A2.6 Direct Products

Let K and G to be two groups. We assume for now that both K and G are multiplicative. The *(external) direct product* of K and G is the group

$$H = K \times G = \{(k, g) : k \in K,\ g \in G\}$$

with componentwise multiplication

$$(k, g)(k', g') = (kk', gg').$$

Note that we may identify K with the subgroup $\{(k, 1) : k \in K\}$, and identify G with the subgroup $\{(1, g) : g \in G\}$. With this identification, we observe that the subgroups K and G are complementary, i.e. $H = KG = \{hk : h \in K,\ k \in G\}$ (recall the identification of k with $(k, 1)$ and g with $(1, g)$) and $K \cap G = 1$ (so that every $g \in H$ can be *uniquely* expressed as $h = kg$ for k and g as above). Moreover these two subgroups are normal and they commute with each other: $kg = gk$ for all $k \in K$ and $g \in G$.

Conversely, given a group H, in order to recognize H as the direct product of two subgroups $K, G \leq H$, we require that $H = KG$, $K \cap G = 1$, and K commutes with G (in particular both K and G are normal subgroups). We then write $H = KG = K \times G$, the *(internal) direct product* of K and G.

This construction is entirely analogous to the construction of (internal and external) direct sums of vector spaces.

A2.7 Semidirect Products

Here we generalize the notion of a direct product. Let K and G be groups, and suppose that G acts on K. In this context (as in Section A2.1) G permutes the set of elements of K; but here we require that the resulting permutations of K are actually automorphisms of K. This means that each $g \in G$ determines a map $K \to K$ denoted by $k \mapsto k^g$ such that

$$(h_1 h_2)^k = h_1^k h_2^k; \quad h^{k_1 k_2} = (h^{k_1})^{k_2}$$

for all $k, k_1, k_2 \in K;\ g, g_1, g_2 \in G$. (Note that the data we are given includes not only groups K and G but also a choice of homomorphism $G \to \mathrm{Aut}(K)$.) Define the *(external) semidirect product* of K and G as

$$H = K \rtimes G = \{(k, g) : k \in K,\ g \in G\}$$

where the product in H is defined by

$$(k_1, g_1)(k_2, g_2) = (k_1^{g_2} k_2, g_1 g_2)$$

for all $k_i \in K,\ g_i \in G$. If you have never done this before, you should check that this actually does define a group; most importantly, this product is associative. Again $\{(k, 1) : k \in K\}$ is a subgroup (actually a normal subgroup) which we identify with K; and $\{(1, g) : g \in G\}$ is a subgroup (although not in general normal) which we identify with G. Note that K and G do not typically commute with each other; indeed

$$(1, g)^{-1}(k, 1)(1, g) = (k^g, 1)$$

so that the original action of G on K which was given, is realized as the action by conjugation in the group H. It is important to realize that the data required to construct the group H includes not only the groups K and G, but also the choice of action of G on K. In particular if one chooses the trivial action, one obtains simply a direct product as a special case.

Reversing our viewpoint, suppose we are given a group H and two subgroups $K, G \leq H$ such that K is normal and every element $h \in H$ is uniquely expressible as $h = kg$ where $k \in K,\ g \in G$ (i.e. $H = KG$ with $K \cap G = 1$). Then H is the *(internal) semidirect product* of K and G.

As an example, the group of affine linear transformations on V is

$$AGL(V) = V \rtimes GL(V)$$

where V is identified as the normal subgroup consisting of translations $v \mapsto v + a$, and $GL(V)$ is identified as the subgroup consisting of linear transformations $v \mapsto Av$. Similarly,

$$A\Gamma L(V) = V \rtimes \Gamma L(V).$$

Also the group of semilinear transformations on V is

$$\Gamma L = GL(V) \rtimes \mathrm{Aut}\, F.$$

As a final example, consider a cyclic group $K = \{1, x, x^2, \ldots, x^{n-1}\}$ of order n, and let $G = \{1, y\}$ be a group of order 2. Then any semidirect product of K by G is either a direct product (in which x commutes with y) or a dihedral group (in which $x^y = y^{-1}xy = x^{-1}$).

Appendix A3: Algebras and Representations

An **algebra** over a field F is a vector space over F which also a ring with identity (and certain assumptions of compatibility between the ring and vector space structures are required to hold, including distributivity of multiplication over addition). Although an identity element $1 \in A$ is not assumed in every case, we will only consider algebras with identity. Formally, we begin with a vector space A over a field F. We assume that A also has a multiplicative structure: $xy \in A$ whenever $x, y \in A$ and that A has two distinguished elements $\mathbf{0}, \mathbf{1} \in A$ such that

$$x + \mathbf{0} = \mathbf{0} + x = \mathbf{1}x = x\mathbf{1} = x$$

for all $x \in A$. If no confusion arises, we write simply $0, 1$ instead of $\mathbf{0}, \mathbf{1}$ (although these elements are distinct from the scalars $0, 1 \in F$). Moreover we assume that the following statements hold for all $r, s \in F$ and $x, y \in A$:

$$(rx)y = r(xy) = x(ry);$$
$$(rs)x = r(sx).$$

In this case we call A an algebra over F. We say A is **commutative** or **non-commutative** according as its *multiplication* is commutative or not (regardless of the addition in A, which is *always* commutative).

A3.1 Example: Matrix Algebras. The set $F^{n\times n}$ consisting of all $n \times n$ matrices over a field F is an algebra over F. This algebra is not commutative when $n \geq 2$. We usually denote $\mathbf{0}$ (the zero matrix) and $\mathbf{1}$ (the identity matrix) simply by 0 and I.

A3.2 Example: Polynomial Algebras. The polynomial ring $R = F[X_1, X_2, \ldots, X_n]$ consisting of all polynomials in $X_1, X_2, \ldots, X_n$ with coefficients in F, is an algebra over F. Here $\mathbf{0}$ and $\mathbf{1}$ are just the constant polynomials 0 and 1. It is a **graded ring**, i.e.

$$R = \bigoplus_{k\geq 0} R_k\,, \qquad R_k R_\ell \subseteq R_{k+\ell}\,.$$

Here R_k is the subspace of all k-homogeneous polynomials. By Lemma 21.6 its dimension is $\binom{n-1+k}{k}$.

A3.3 Example: Group Algebras. Let G be a finite multiplicative group, and let F be a field. The **group algebra** of G over F is the set FG consisting of all formal linear combinations of elements of G with coefficients in F. Typical elements $a, b \in FG$ have the form

$$a = \sum_{g\in G} a_g g; \qquad b = \sum_{g\in G} b_g g$$

where $a_g, b_g \in F$; addition and multiplication in FG are then defined by

$$a+b=\sum_{g\in G}(a_g+b_g)g; \qquad ab=\sum_{g,h\in G}a_gb_hgh=\sum_{x\in G}\Big(\sum_{g\in G}a_gb_{g^{-1}x}\Big)x.$$

Note that the group algebra FG is commutative iff the group G is itself abelian.

For example consider $G=S_3=\{(1),(12),(13),(23),(123),(132)\}$. In this case

$$FS_3=\{a(1)+b(12)+c(13)+d(23)+e(123)+f(132)\,:\,a,b,c,d,e,f\in F\}.$$

An example of multiplication in this algebra is given by

$$\big[(12)-2(13)\big]\big[(1)+(12)-(23)+(123)\big]=(1)+(12)-(13)-2(23)+2(123)-3(132).$$

Here the zero element $\mathbf{0}\in FG$ is the linear combination of group elements in which all coefficients are zero; it is usually denoted simply 0. Also the algebra identity $\mathbf{1}\in FG$ is just the group identity, usually denoted simply by 1, or in this case (1).

A **skewfield** (or **division ring**[1]) is a ring with identity, in which every nonzero element is a unit (i.e. is invertible). A **field** is a commutative skewfield. A **division algebra** over a field F is an algebra over F which is a skewfield (i.e. division ring).

A3.4 Theorem (Wedderburn). Every finite skewfield is a field.

A3.5 Example: The Real Quaternions. The algebra of real quaternions is the 4-dimensional algebra over $\mathbb{R}$ given by

$$\mathbb{H}=\{a+bi+cj+dk\,:\,a,b,c,d\in\mathbb{R}\}$$

where

$$i^2=j^2=k^2=ijk=-1;$$
$$ij=k; \quad jk=i; \quad ki=j;$$
$$ji=-k; \quad kj=-i; \quad ik=-j.$$

This is the easiest example of a non-commutative skewfield.

Let A be an algebra over F. A **(left) A-module** is a vector space M over F together with a left action of A on M, written as left-multiplication ($ax\in M$ whenever $a\in A$, $x\in M$) such that

$a(x+y)=ax+ay$;

[1] Caution: The term *division ring* is sometimes used in a different sense by finite geometers; see [31].

$(a+b)x = ax + bx$;
$a(bx) = (ab)x$;
$\mathbf{1}x = x$

for all $a, b \in A$; $x, y \in M$. Given an A-module M, an F-subspace $N \leq M$ is an **A-submodule** if $ax \in N$ whenever $a \in A$, $x \in N$.

A3.6 Example: The Natural Module for $F^{n\times n}$. Let $F^{n\times n}$ be the algebra of $n \times n$ matrices over a field F. Then F^n (the n-dimensional vector space consisting of column vectors of length n over F) is a module for $F^{n\times n}$, in which matrices act on the left by the usual matrix multiplication.

A3.7 Example: The Regular Module. Let A be an algebra over F. Then A may be regarded as a module over itself, where the left action of A on itself is the usual multiplication in A. This module is called the **(left) regular A-module**. Note that a submodule of the left regular module, is the same thing as a left ideal of A.

Let A be an algebra over F, and let M be a nonzero A-module. We say M is **simple** (or **minimal** or **irreducible**) if its only A-submodules are (0) and M itself. We say M is **semisimple** if it satisfies the equivalent conditions (i) and (ii) of the following.

A3.8 Proposition. Let A be an algebra over F, and let M be a finite-dimensional A-module. Then the following two conditions are equivalent.

(i) For every submodule $U \subseteq M$ there exists a complementary ideal $U' \subseteq M$ with $M = U \oplus U'$.

(ii) There exist simple submodules $U_1, U_2, \ldots, U_r \subseteq M$ such that $M = U_1 \oplus U_2 \oplus \cdots \oplus U_r$.

A3.9 Example: Modules for $F^{n\times n}$. It is easy to see that the natural module F^n for $F^{n\times n}$ is simple. The regular module is semisimple:

$$F^{n\times n} \cong U_1 \oplus U_2 \oplus \cdots \oplus U_n$$

where $U_i \subseteq F^{n\times n}$ is the submodule (i.e. left ideal) consisting of all $n \times n$ matrices having zeroes outside of the i-th column.

Let G be a finite multiplicative group and F a field. A **(linear) representation** of G over F is a homomorphism $\pi : G \to GL_n(F)$. Given such a representation, the vector space F^n has the structure of a module over the group algebra FG, where a typical element $g \in G$ acts on $v \in F^n$ as $v \mapsto \pi(g)v$, and more generally

$$a = \sum_{g\in G} a_g g \in FG$$

acts on F^n as

$$v \mapsto \sum_{g\in G} a_g \pi(g) v\,.$$

Conversely, an action of the group algebra FG on F^n gives rise to a representation $G \to GL_n(F)$, so the two notions of a representation of G, and a module for FG, are really equivalent. A useful sufficient condition for semisimplicity of FG-modules is the following.

A3.10 Theorem (Maschke). Let G be a finite group, and let F be a field whose characteristic does not divide $|G|$. (Thus either $\operatorname{char} F = 0$ or $\operatorname{char} F = p \not| |G|$.) Then every FG-module is semisimple.

Proof. Let M be an FG-module and let $U \leq M$ be a submodule. Choose any complementary subspace $W \leq M$, so that $M = U \oplus W$ as F-vector spaces. We cannot assume that W is an FG-submodule; if it were, we would be done. Let $P : M \to U$ be the projection onto U along W; thus $P : M \to U$ is the unique F-linear transformation satisfying

$Px = x$ iff $x \in U$;
$Px = 0$ iff $x \in W$;
$P^2 = P$.

Define the F-linear transformation

$$\widetilde{P} : M \to U; \qquad v \mapsto \frac{1}{|G|} \sum_{h\in G} h^{-1} P h v.$$

Note that the scalar $|G|$ is invertible in F, by the hypothesis on the characteristic of F. We first show that

(A3.11) $\qquad \widetilde{P}(av) = a(\widetilde{P}v)$ for all $a \in FG,\ v \in V$.

It suffices to verify (A3.11) in the case $a = g \in G$; but in this case

$$\begin{aligned}
\widetilde{P}(gv) &= \frac{1}{|G|} \sum_{h\in G} h^{-1} P h g v \\
&= \frac{1}{|G|} \sum_{k\in G} g k^{-1} P k v \qquad \text{where } k = hg \\
&= \frac{1}{|G|} g \sum_{k\in G} k^{-1} P k v \\
&= g(\widetilde{P}v)
\end{aligned}$$

and so (A3.11) follows in the general case from linearity. Next we check that

(A3.12) $\widetilde{P}^2 = \widetilde{P}$.

Indeed, for all $v \in M$ we have

$$\begin{aligned}\widetilde{P}^2 v &= \frac{1}{|G|^2} \sum_{h,k\in G} h^{-1} P h k^{-1} P k v \\ &= \frac{1}{|G|^2} \sum_{h,k\in G} h^{-1} h k^{-1} P k v\end{aligned}$$

since $Pkv \in U$ implies that $hk^{-1}Pkv \in U$ also, and so the latter element is fixed by P. Thus

$$\begin{aligned}\widetilde{P}^2 v &= \frac{1}{|G|^2} \sum_{h,k\in G} k^{-1} P k v \\ &= \frac{1}{|G|^2} \sum_{k\in G} |G| k^{-1} P k v \\ &= \frac{1}{|G|} \sum_{k\in G} k^{-1} P k v \\ &= \widetilde{P} v\end{aligned}$$

which proves (A3.11). Thus $\widetilde{P}$ is a projection operator. We show that it is in fact a projection onto U:

(A3.12) $\widetilde{P}v = v$ iff $v \in U$.

First observe that if $v \in U$ then

$$\begin{aligned}\widetilde{P} v &= \frac{1}{|G|} \sum_{h\in G} h^{-1} P h v \\ &= \frac{1}{|G|} \sum_{h\in G} h^{-1} h v\end{aligned}$$

since $hv \in U$ is fixed by P; thus $\widetilde{P}v = v$. Conversely, suppose that $\widetilde{P}v = v$; then since $\widetilde{P}v \in U$ it follows that $v \in U$. Thus (A3.12) holds. Finally

(A3.12) $M = U \oplus U'$ where $U' = \ker P$ is an FG-submodule.

If $v \in U' = \ker P$ then for all $a \in FG$ we have

$$P(av) = a(Pv) = 0$$

by (A3.11), so $av \in U'$ as well. Thus $U' \subseteq M$ is an FG-submodule, and the remaining assertions in (A3.12) follow from elementary properties of projections. □

A3.13 Corollary. Let G be a finite group, and let F be a field of order not dividing $|G|$. (Thus either $\operatorname{char} F = 0$ or $\operatorname{char} F = p \not| \, |G|$.) Then for every left ideal $\mathfrak{I} \subseteq FG$ there is a complementary left ideal $\mathfrak{I}' \subseteq FG$ satisfying $FG = \mathfrak{I} \oplus \mathfrak{I}'$.

Proof. Apply Theorem A3.10 to the regular module FG. □

Let A be an algebra over F. Elements of the centre $Z(A)$ are called **central** elements of A. An **idempotent** in A is an element $e \in A$ such that $e^2 = e$.

Now suppose that A is commutative and that the regular module of A is semisimple. We exhibit a one-to-one correspondence[1] between the idempotents of A, and the ideals of A. First suppose $\mathfrak{I} \subseteq A$ is an ideal. By assumption there exists a complementary ideal $\mathfrak{I}' \subseteq A$ such that $A = \mathfrak{I} \oplus \mathfrak{I}'$. (Note that all ideals are two-sided since A is commutative.) In particular

$$\mathbf{1} = e + e' \tag{A3.14}$$

for unique elements $e \in \mathfrak{I}$, $e' \in \mathfrak{I}'$. From (A3.14) we obtain $e = e^2 + e'e$ where $e'e \in \mathfrak{I} \cap \mathfrak{I}' = (0)$, so $e^2 = e$ and similarly $(e')^2 = e'$. Right-multiplying (A3.14) by an arbitrary $x \in A$ gives

$$x = ex + e'x \tag{A3.15}$$

where $ex \in \mathfrak{I}$ and $e'x \in \mathfrak{I}'$; thus the map $x \mapsto ex$ is the projection of A onto $\mathfrak{I}$ along $\mathfrak{I}'$; similarly $x \mapsto e'x$ is the projection of A onto $\mathfrak{I}'$ along $\mathfrak{I}$. Thus $\mathfrak{I} = Ae$ and $\mathfrak{I}' = Ae'$. Evidently e is the unique idempotent generator of $\mathfrak{I}$ so the correspondence $\mathfrak{I} \leftrightarrow e$ is bijective. The essential features of this correspondence are summarized as follows.

A3.16 Theorem. Let A be a commutative algebra over F whose regular module is semisimple. Then there is a one-to-one correspondence between idempotent elements $e \in A$ and ideals $\mathfrak{I} \subseteq A$, such that $\mathfrak{I} = Ae$. Moreover if $\mathfrak{I}_1 = Ae_1$ and $\mathfrak{I}_2 = Ae_2$ where each e_i is idempotent, then

(i) $\mathfrak{I}_1\mathfrak{I}_2 = \mathfrak{I}_1 \cap \mathfrak{I}_2 = Ae_1e_2$ and

(ii) $\mathfrak{I}_1 + \mathfrak{I}_2 = A(e_1 + e_2 - e_1e_2)$

where e_1e_2 and $e_1 + e_2 - e_1e_2$ are the unique idempotent generators of the ideals listed in (i) and (ii) respectively.

[1] A similar conclusion holds under the weaker hypothesis that A is semisimple, meaning that its Jacobson radical is (0), but A is not necessarily commutative. In this case there is a one-to-one correspondence between the (two-sided) ideals of A and the *central* idempotents of A; see e.g. [32]. We will not require this stronger version.

Proof. We have already established the indicated one-to-one correspondence. Consider ideals $\mathfrak{I}_i = Ae_i$ for $i = 1, 2$ where e_1 and e_2 are idempotent. Then

$$(e_1e_2)^2 \;=\; e_1^2e_2^2 \;=\; e_1e_2;$$
$$(e_1{+}e_2{-}e_1e_2)^2 \;=\; e_1^2{+}e_2^2{+}e_1^2e_2^2{+}2e_1e_2{-}2e_1^2e_2{-}2e_1e_2^2 \;=\; e_1{+}e_2{-}e_1e_2$$

so these two elements are idempotent. Clearly

$$\mathfrak{I}_1 \cap \mathfrak{I}_2 \supseteq \mathfrak{I}_1\mathfrak{I}_1 = Ae_1e_2\,.$$

But if $x \in \mathfrak{I}_1 \cap \mathfrak{I}_2$ then $x = e_1x = e_1e_2x \in Ae_1e_2$, so equality holds in (i). For (ii) it is clear that

$$\mathfrak{I}_1 + \mathfrak{I}_2 \supseteq A(e_1{+}e_2{-}e_1e_2);$$

to check equality, note that for all $x \in A$ we have

$$(xe_1)(e_1{+}e_2{-}e_1e_2) \;=\; x(e_1^2{+}e_1e_2{-}e_1^2e_2) \;=\; xe_1$$

so that $\mathfrak{I}_1 \subseteq A(e_1{+}e_2{-}e_1e_2)$. A similar argument shows that $\mathfrak{I}_2 \subseteq A(e_1{+}e_2{-}e_1e_2)$, so in fact

$$\mathfrak{I}_1 + \mathfrak{I}_2 \subseteq A(e_1{+}e_2{-}e_1e_2)$$

also, and so the equality (ii) holds. □

A3.17 Example: Diagonal Matrices. Let A be the ring of $n \times n$ diagonal matrices over F. [Of course $A \cong F^n = F \oplus F \oplus \cdots \oplus F$ (n summands).] Then A has 2^n ideals, corresponding to the 2^n idempotent elements, these being the diagonal matrices with diagonal entries $\in \{0, 1\}$. To be even more concrete, let's take $n = 4$ and consider the ideals

$$\mathfrak{I}_1 = Ae_1 = \left\{ \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \;:\; a, b \in F \right\} \quad \text{where } \; e_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix};$$

$$\mathfrak{I}_2 = Ae_2 = \left\{ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \;:\; a, b \in F \right\} \quad \text{where } \; e_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then

$$\mathfrak{I}_1\mathfrak{I}_2 = \mathfrak{I}_1 \cap \mathfrak{I}_2 = Ae_1e_2 = \left\{ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \;:\; a \in F \right\} \quad \text{where } \; e_1e_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$\mathfrak{I}_1{+}\mathfrak{I}_2 = A(e_1{+}e_2{-}e_1e_2) = \left\{ \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \;:\; a, b, c \in F \right\}$$
$$\text{where } \; e_1{+}e_2{-}e_1e_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Appendix A4: Exterior Algebra

Let V be an n-dimensional vector space over F with basis $e_1, e_2, \ldots, e_n$. The **k-th exterior power** of V is the vector space $\bigwedge^k V$ of dimension $\binom{n}{k}$ with basis given by the symbols

$$e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_k}\,, \qquad 1 \le i_1 < i_2 < \cdots < i_k \le n.$$

(The symbol '$\wedge$' is read as 'wedge'.) We extend this notation by defining $v_1 \wedge v_2 \wedge \cdots \wedge v_k \in \bigwedge^k V$ for all $v_1, v_2, \ldots, v_k \in V$ subject to the rules:

(A4.1) The vector $v_1 \wedge v_2 \wedge \cdots \wedge v_k$ is linear in each argument if the other arguments are fixed, i.e.

$$\begin{aligned} &v_1 \wedge v_2 \wedge \cdots \wedge (av_i + bv_i') \wedge \cdots \wedge v_k \\ &\qquad = a(v_1 \wedge v_2 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_k) \;+\; b(v_1 \wedge v_2 \wedge \cdots \wedge v_i' \wedge \cdots \wedge v_k)\,. \end{aligned}$$

(A4.2) The vector $v_1 \wedge v_2 \wedge \cdots \wedge v_k$ vanishes if any two of its arguments coincide. Thus if $v_i = v_j$ for some $i \ne j$ then $v_1 \wedge v_2 \wedge \cdots \wedge v_k = 0$.

From (A4.1) and (A4.2) we obtain

(A4.3) The vector $v_1 \wedge v_2 \wedge \cdots \wedge v_k$ is replaced by its negative, if any two of its arguments are interchanged. Thus for all $i \ne j$ we have

$$v_1 \wedge v_2 \wedge \cdots \wedge v_j \wedge \cdots \wedge v_i \wedge \cdots \wedge v_k = -(v_1 \wedge v_2 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_j \wedge \cdots \wedge v_k).$$

The identity (A4.3) follows from (A4.1) and (A4.2) by simply expanding

$$\begin{aligned} 0 = {} & v_1 \wedge v_2 \wedge \cdots \wedge (v_i + v_j) \wedge \cdots \wedge (v_i + v_j) \wedge \cdots \wedge v_k \\ & - (v_1 \wedge v_2 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_i \wedge \cdots \wedge v_k) \;-\; (v_1 \wedge v_2 \wedge \cdots \wedge v_j \wedge \cdots \wedge v_j \wedge \cdots \wedge v_k). \end{aligned}$$

Conversely we may derive (A4.2) from (A4.1) and (A4.3), assuming $\operatorname{char} F \ne 2$. (If $\operatorname{char} F = 2$ then nothing can be deduced from (A4.3) since the assertion (A4.3) says merely that a vector equals itself.) The properties above suffice to evaluate an arbitrary wedge product $v_1 \wedge v_2 \wedge \cdots \wedge v_k$; however if one requires a more explicit definition then the following will serve. First expand $v_i = \sum_{1 \le j \le n} v_{ij} e_j$ where $v_{ij} \in F$; then

$$\text{(A4.4)} \qquad v_1 \wedge v_2 \wedge \cdots \wedge v_k \;=\; \sum_{1 \le j_1 < j_2 < \cdots < j_k \le n} \begin{vmatrix} v_{1j_1} & v_{1j_2} & \cdots & v_{1j_k} \\ v_{2j_1} & v_{2j_2} & \cdots & v_{2j_k} \\ \vdots & \vdots & \ddots & \vdots \\ v_{kj_1} & v_{kj_2} & \cdots & v_{kj_k} \end{vmatrix} e_{j_1} \wedge e_{j_2} \wedge \cdots \wedge e_{j_k}\,.$$

Consider the matrix M whose rows are the coordinates of $v_1, v_2, \ldots, v_k$ with respect to $e_1, e_2, \ldots, e_n$. This $k \times n$ matrix M has $\binom{n}{k}$ submatrices of size $k \times k$, whose determinants are simply the coefficients in (A4.4). Recall that M is singular iff all of its $k \times k$ submatrices are singular; thus

(A4.5) Vectors $v_1, v_2, \ldots, v_k \in V$ are linearly dependent iff $v_1 \wedge v_2 \wedge \cdots \wedge v_k = 0$.

We naturally identify $\bigwedge^1 V = V$; also $\bigwedge^0 V = F$. If $v \in \bigwedge^k V$ and $w \in \bigwedge^\ell V$ then $v \wedge w \in \bigwedge^{k+\ell} V$ is well-defined; here one simply identifies

$$(v_1 \wedge v_2 \wedge \cdots \wedge v_k) \wedge (w_1 \wedge w_2 \wedge \cdots \wedge w_\ell) = v_1 \wedge v_2 \wedge \cdots \wedge v_k \wedge w_1 \wedge w_2 \wedge \cdots \wedge w_\ell \,.$$

When $k = 0$ or $\ell = 0$, the corresponding factor v or w is an empty wedge product, which by definition is simply the scalar 1. Also it follows from the definitions that $\bigwedge^k V = 0$ whenever $k > n$.

The **exterior algebra** of V is the algebra

$$\bigwedge V = \bigoplus_{k \geq 0} \bigwedge^k V = F \oplus V \oplus \bigwedge^2 V \oplus \bigwedge^3 V \oplus \cdots \oplus \bigwedge^n V$$

with respect to the wedge product '$\wedge$'. This is a graded algebra of dimension

$$\sum_{0 \leq k \leq n} \binom{n}{k} = 2^n.$$

By 'graded' we mean simply the fact that multiplying (i.e. 'wedging') elements of $\bigwedge^k V$ with elements of $\bigwedge^\ell V$, gives elements of $\bigwedge^{k+\ell} V$, as previously mentioned.

The inexperienced reader is encouraged to think of V as F^n, and to identify $\bigwedge^2 V$ with the $\binom{n}{2}$-dimensional space $\mathrm{Skew}_n(F)$ of all skew-symmetric $n \times n$ matrices over F. An actual isomorphism is given by

$$\bigwedge^2 V \to \mathrm{Skew}_n(F), \qquad v \wedge w \mapsto v^T w - w^T v.$$

Note that an $n \times n$ matrix of the form $v^T w$ has rank 1 (unless $v = w = 0$, in which case $v^T w = 0$ has rank 0) and that the skew-symmetric matrix $v^T w - w^T v$ has rank either 0 or 2, according as v and w are or are not linearly dependent. In general, a skew-symmetric $n \times n$ matrix is not necessarily of the form $v^T w - w^T v$; elements of $\mathrm{Skew}_n(F)$ can have any rank in $0, 2, 4, \ldots, 2k$ where $n \in \{2k, 2k+1\}$. However, $\mathrm{Skew}_n(F)$ is *spanned by* its rank 2 elements.

The preceding remarks are intended to help the reader remember that not every vector in $\bigwedge^k V$ has the form 'pure' form $v_1 \wedge v_2 \wedge \cdots \wedge v_k$ where $v_1, v_2, \ldots, v_k \in V$; rather, $\bigwedge^k V$ is spanned by elements of this pure form.

Appendix A5: Coding Theory

Coding theory, or the theory of error-correcting codes, concerns how information may best be represented for the purpose of transmission over a noisy channel, or stored in imperfect media, in such a way that a limited number of errors introduced during transmission/reception, or storage/retrieval, may be corrected. To formalize these goals inevitably leads us to notions of finite geometry. Coding theorists use many of the standard constructions of finite geometry in the construction of good codes. In this study, where finite geometry is our primary interest, we shall have more use for the reverse process in which coding theory is used to study finite geometry. For a more comprehensive introduction to coding theory, see e.g. [65].

An **alphabet** is simply a finite set A of symbols, possibly the Roman alphabet $\{a, b, c, \ldots, z\}$ or the set of 256 ASCII characters; or more typically for our use, the elements of a finite field, especially the **binary alphabet** $\mathbb{F}_2 = \{0, 1\}$ whose elements we call **bits**. A **code** of length n over an alphabet A is a subset $\mathcal{C} \subseteq A^n$. We refer to $\mathcal{C}$ as a **q-ary code** where $q = |A|$. Elements of A^n are called **words** (or **bitstrings** if $A = \{0, 1\}$), and elements of the subset $\mathcal{C}$ are called **codewords**. The **(Hamming) distance** between two words $w, w' \in A^n$, denoted $\delta(w, w') \in \{0, 1, 2, \ldots, n\}$, is the number of coordinates in which they differ. Note that δ is a metric on A^n. The **minimum distance** of $\mathcal{C}$ is

$$d \;=\; \min_{\substack{w \neq v \\ \text{in } \mathcal{C}}} \delta(v, w).$$

Denote the **closed ball of radius** r centred at a vector $w \in \mathbb{F}_q^n$ by

$$B_r(w) = \{v \in \mathbb{F}_q^n \,:\, \delta(v, w) \leq r\}.$$

A code $\mathcal{C}$ is **e-error correcting** if the balls $B_e(w)$ centred at the codewords $w \in \mathcal{C}$ are mutually disjoint. In this case a received word $v \in B_e(w)$ is uniquely decoded as $w \in \mathcal{C}$, at least in principle. (Actually finding the closest codeword $w \in \mathcal{C}$ to a given word v may be difficult in practice.)

A5.1 Proposition. A code $\mathcal{C}$ is e-error correcting iff its minimum distance is at least $2e + 1$.

Proof. If $\mathcal{C}$ is not e-error correcting then there exist $w \neq w'$ in $\mathcal{C}$ such that $B_e(w) \cap B_e(w')$ contains some vector $v \in \mathbb{F}_q^n$; but then $\delta(w, w') \leq \delta(w, v) + \delta(v, w') \leq e + e = 2e$. Clearly the converse of this argument also holds. □

We say that $\mathcal{C}$ is **e-error detecting** if $B_e(w) \cap B_{e-1}(w') = \emptyset$ for all $w \neq w'$ in $\mathcal{C}$. This says that if at most e symbols in a transmitted word are altered during transmission, the recipient can be sure that such an error occurred; but cannot in general correct it; see

Example A5.6 below. Clearly every code with minimum distance d is $\lfloor \frac{d}{2} \rfloor$-error detecting, and $\lfloor \frac{d-1}{2} \rfloor$-error correcting.

Note that

$$|B_r(v)| = 1 + n(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{r}(q-1)^r.$$

An elementary counting argument using this yields

A5.2 Theorem (Sphere-Packing Bound; Hamming Bound). If $\mathcal{C}$ is an e-error correcting q-ary code of length n, then

$$|\mathcal{C}| \leq \frac{q^n}{1 + n(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{r}(q-1)^r}.$$

A **perfect code** is one for which equality holds in the Sphere-Packing Bound; in this case the balls $B_e(w)$ centred at codewords $w \in \mathcal{C}$ partition the set A^n of all words.

Usually $A = \mathbb{F}_q$ is a finite field, and subspaces $\mathcal{C} \leq \mathbb{F}_q^n$ are called **linear codes**. A linear code of length n and dimension k is called an $[n,k]$**-code**. The **(Hamming) weight** of a word (i.e. vector) $v \in \mathbb{F}_q^n$ is the number $wt(v) \in \{0, 1, 2, \ldots, n\}$ of nonzero coordinates of v. Note that the Hamming distance between two words $v, w \in \mathbb{F}_q^n$ is given by

$$\delta(v, w) = wt(w - v).$$

It follows that the minimum distance of a linear code $\mathcal{C}$ coincides with the **minimum weight**:

$$d = \min_{\substack{w \neq 0 \\ \text{in } \mathcal{C}}} wt(w).$$

A linear code with minimum weight d is also called an $[n,k,d]$**-code**.

Let G be a $k \times n$ matrix over $\mathbb{F}_q$ having linearly independent rows (so in particular $k \leq n$). The row space of G is an $[n,k]$-code

$$\mathcal{C} = \{xG : x \in \mathbb{F}_q^k\}.$$

We call G a **generator matrix** for the code $\mathcal{C}$. In this case the injective map $\mathbb{F}_q^k \mapsto \mathbb{F}_q^n$, $x \mapsto xG$ gives an encoding of all q-ary words of length k, as words of length n. We say that $\mathcal{C}$ has **information rate** equal to $\frac{k}{n}$. This is the fraction of the transmitted symbols which contain useful information; the remaining symbols may serve to provide error correction.

An $[n,k]$-code may be specified as the row space of a $k \times n$ matrix G as above, or as the right null space of an $(n-k) \times n$ matrix H having linearly independent rows:

$$\mathcal{C} = \{y \in \mathbb{F}_q^n \,:\, Hy^T = 0 \in \mathbb{F}_q^{n-k}\}.$$

Such a matrix H is called a **parity check matrix** for $\mathcal{C}$. Given a received word $y \in \mathbb{F}_q^n$, we call the vector $Hy^T \in \mathbb{F}_q^{n-k}$ the **(error) syndrome** of y. The syndrome is useful in error detection and correction: certainly a nonzero syndrome indicates that an error occurred during transmission. To see the utility of syndromes in error correction, see Example A5.8 below.

The fundamental goal of coding theory is to find codes with the following desirable properties:

- High information rate;
- High error-correcting capability (i.e. large minimum distance);
- Efficient encoding and decoding algorithms should be available.

These goals tend to be conflicting in nature; for example one can increase the error-correcting capability, at the cost of reducing the information rate and thereby increasing the cost of transmission, as Theorem A5.2 shows. Another bound which shows this trade-off between minimum distance and information rate is the following.

A5.3 Theorem (Singleton Bound). If $\mathcal{C}$ is a linear $[n,k,d]$-code over $\mathbb{F}_q$ then $k \leq n - d + 1$.

Proof. Let H be an $(n-k) \times n$ parity check matrix for $\mathcal{C}$. Since H has rank $n-k$, it has a set of $n-k$ linearly independent columns; we may suppose that the first $n-k$ columns of H are linearly independent, otherwise permute the columns of H appropriately. (Permuting the coordinates of $\mathcal{C}$ yields once again an $[n,k,d]$-code.) Now we may assume H is in reduced row-echelon form; otherwise replace H by its reduced row-echelon form, and this will not change its right null space. Thus

$$H = \begin{bmatrix} I_{n-k} & X \end{bmatrix}$$

where I is the $(n-k) \times (n-k)$ identity matrix, and $X \in \mathbb{F}_q^{(n-k)\times k}$. It is easy to check that the matrix

$$G = \begin{bmatrix} -X^T & I_k \end{bmatrix}$$

is a generator matrix for $\mathcal{C}$. Its rows have weight at most $n-k+1$, so that $d \leq n-k+1$. □

Codes for which the Singleton Bound is attained, are called **MDS codes**. (This is an acronym for *Maximum Distance Separable,* which is such an unfortunate term that

everyone just calls them MDS codes.) Although both bounds A5.2 and A5.3 describe the optimal possible tradeoff between information rate and minimum distance, neither bound implies the other. Thus there exist MDS codes which are not perfect; and there exist perfect codes which are not MDS.

The **weight enumerator** of a code $\mathcal{C}$ of length n is the polynomial

$$A_{\mathcal{C}}(z) = \sum_{v \in \mathcal{C}} x^{n-wt(v)} y^{wt(v)} = \sum_{0 \le d \le n} A_d x^{n-d} y^d$$

where A_d is the number of codewords of weight d. The essential information conveyed by the weight enumerator is the list of integer values $A_0, A_1, \ldots, A_n$, known as the **weight distribution** of $\mathcal{C}$. The reason for representing this list of values as a single polynomial is motivated by Theorem A5.4 below. In the case $\mathcal{C}$ is a linear code, we define the **dual code** by

$$\mathcal{C}^{\perp} = \{w \in \mathbb{F}_q^n \ : \ w \cdot v = 0 \text{ for all } v \in \mathcal{C}\}$$

where $w \cdot v = wv^T$. Thus the dual of an $[n,k]$-code is an $[n, n-k]$-code. If $\mathcal{C}$ has generator matrix G and parity check matrix H, then $\mathcal{C}^{\perp}$ has generator matrix H and parity check matrix G.

The MacWilliams relations (Theorem A5.4) show that the weight distribution of either of these two codes ($\mathcal{C}$ or $\mathcal{C}^{\perp}$) determines the weight distribution of the other.

A5.4 Theorem (MacWilliams). Let $\mathcal{C}$ be a linear $[n,k]$-code over $\mathbb{F}_q$. Then the weight enumerator of the dual code is given by

$$A_{\mathcal{C}^{\perp}}(x,y) = q^{-k} A_{\mathcal{C}}(x + (q-1)y,\ x - y).$$

Before proving Theorem A5.4 we consider an example.

A5.5 Example. Consider the binary $[5,2]$-code $\mathcal{C} = \{00000, 11100, 10011, 01111\}$, which has weight enumerator

$$A_{\mathcal{C}}(x,y) = x^5 + 2x^2y^3 + xy^4.$$

Its dual is the binary $[5,3]$-code

$$\mathcal{C}^{\perp} = \{00000, 01100, 00011, 01111, 11010, 11001, 10110, 10101\}$$

whose weight enumerator is

$$A_{\mathcal{C}^{\perp}}(z) = x^5 + 2x^3y^2 + 4x^2y^3 + xy^4.$$

The MacWilliams relations are expressed as the identity

$$A_{\mathcal{C}^\perp}(x,y) = \tfrac{1}{4} A_{\mathcal{C}}(x+y,\, x-y),$$

i.e.

$$x^5 + 2x^3y^2 + 4x^2y^3 + xy^4 = \tfrac{1}{4}\big[(x+y)^5 + 2(x+y)^2(x-y)^3 + (x+y)(x-y)^4\big]$$

which may be verified directly. In passing we note that $\mathcal{C}^\perp \cap \mathcal{C} = \{00000, 01111\}$. This is typical in that while the subspaces $\mathcal{C}, \mathcal{C}^\perp \leq \mathbb{F}_q^n$ have complementary dimension, they are not in general complementary subspaces (unlike the situation for inner product spaces).

Proof of Theorem A5.4. A **character** of the additive group of $\mathbb{F}_q$ is a map $\chi : \mathbb{F}_q \to \mathbb{C}^\times$ such that $\chi(a+b) = \chi(a)\chi(b)$ for all $a, b \in \mathbb{F}_q$. (Here $\mathbb{C}^\times$ is the multiplicative group of nonzero complex numbers.) In particular the map $\chi(a) = 1$ for all $a \in \mathbb{F}_q$ is the **principal character**; every other character of $\mathbb{F}_q$ is **nonprincipal**. Let χ be a nonprincipal character of $\mathbb{F}_q$ and observe that

$$\sum_{a \in \mathbb{F}_q} \chi(a) = 0.$$

For all $u \in \mathbb{F}_q^n$ define

$$g(u) = \sum_{v \in \mathbb{F}_q^n} \chi(u \cdot v) x^{n - wt(v)} y^{wt(v)}.$$

Then

$$\sum_{u \in \mathcal{C}} g(u) = \sum_{u \in \mathcal{C}} \sum_{v \in \mathbb{F}_q^n} \chi(u \cdot v) x^{n-wt(v)} y^{wt(v)} = \sum_{v \in \mathbb{F}_q^n} \Big(\sum_{u \in \mathcal{C}} \chi(u \cdot v) \Big) x^{n-wt(v)} y^{wt(v)}.$$

The innermost sum is q^k for $v \in \mathcal{C}^\perp$ and vanishes otherwise; therefore

$$\sum_{u \in \mathcal{C}} g(u) = q^k \sum_{v \in \mathcal{C}^\perp} x^{n-wt(v)} y^{wt(v)} = q^k A_{\mathcal{C}^\perp}(x,y).$$

For $v = (v_1, \ldots, v_n) \in \mathbb{F}_q^n$ we have

$$wt(v) = wt(v_1) + \cdots + wt(v_n) \quad \text{where } wt(v_i) = \begin{cases} 0, & \text{if } v_i = 0; \\ 1, & \text{if } v_i = 1 \end{cases}$$

and so

$$\begin{aligned} g(u) &= \sum_{v_1, \ldots, v_n \in \mathbb{F}_q} \chi(u_1v_1 + \cdots + u_nv_n) x^{n - wt(v_1) - wt(v_2) - \cdots - wt(v_n)} y^{wt(v_1) + \cdots + wt(v_n)} \\ &= \sum_{v_1, \ldots, v_n \in \mathbb{F}_q} \chi(u_1v_1) x^{1-wt(v_1)} y^{wt(v_1)} \chi(u_2v_2) x^{1-wt(v_2)} y^{wt(v_2)} \cdots \chi(u_nv_n) x^{1-wt(v_n)} y^{wt(v_n)} \\ &= \prod_{1 \leq i \leq n} \sum_{v_i \in \mathbb{F}_q} \chi(u_iv_i) x^{1-wt(v_i)} y^{wt(v_i)}. \end{aligned}$$

The innermost sum equals $x + (q-1)y$ if $u_i = 0$, and $x - y$ if $u_i \neq 0$; thus

$$g(u) = \big(x + (q-1)y\big)^{n-wt(u)}(x-y)^{wt(u)}.$$

Summing over $u \in \mathcal{C}$ gives the required result. □

Consider the problem of transmitting a single hexadecimal digit, or equivalently, a bitstring of length 4. Such a message may be transmitted as a sequence of four bits, but this scheme will provide no error-correcting capability: if the message 0100 is transmitted, and the third bit is altered due to static so that the word 0110 is received, the error cannot be detected by the receiver. The following examples show how we can do better.

A5.6 Example: Parity Check Code. A slightly better scheme would be to add a single parity check bit to every word; thus message words and the corresponding codewords are as listed in the following table. This code detects single errors but corrects no errors.

Message Word	Codeword	Message Word	Codeword
0000	0000 0	1000	1000 1
0001	0001 1	1001	1001 0
0010	0010 1	1010	1010 0
0011	0011 0	1011	1011 1
...	...	...	...
0111	0111 1	1111	1111 0

This code has an information rate of $\frac{4}{5} = 80\%$, meaning that 4 out of every 5 bits transmitted carry useful information; 1 out of every 5 bits transmitted supply redundancy useful in checking the validity of the information transmitted.

This code has generator matrix and parity check matrix given by

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \qquad H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

A5.7 Example: 3-Repetition Code. The idea behind this code is to simply send every bit three times, and to require the receiver to use a simple 'majority rules' approach to decoding. This code is not only 1-error detecting, but in fact 1-error correcting. This

advantage comes at the cost: 2 out of every 3 bits transmitted are redundant bits, and the information rate is only $\frac{1}{3} \approx 33\%$.

Message Word	Codeword	Message Word	Codeword
0000	000 000 000 000	1000	111 000 000 000
0001	000 000 000 111	1001	111 000 000 111
0010	000 000 111 000	1010	111 000 111 000
0011	000 000 111 111	1011	111 000 111 111
...	...	...	...
0111	000 111 111 111	1111	111 111 111 111

This code has generator matrix and parity check matrix given by

$$G = \begin{bmatrix} 1\,1\,1 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 1\,1\,1 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 1\,1\,1 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,1\,1 \end{bmatrix}, \qquad H = \begin{bmatrix} 1\,0\,1 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,1\,1 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 1\,0\,1 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,1\,1 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 1\,0\,1 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,1\,1 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,0\,1 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,1\,1 \end{bmatrix}.$$

A5.8 Example: The $[7, 4, 3]$ binary Hamming Code. The following scheme is superior to the previous example in that it also allows single bit errors to be corrected, yet it has a substantially higher information rate of $\frac{4}{7} \approx 57\%$. It is in fact a perfect code, although not an MDS code.

Message Word	Codeword	Message Word	Codeword
0000	0000 000	1000	1000 011
0001	0001 111	1001	1001 100
0010	0010 110	1010	1010 101
0011	0011 001	1011	1011 010
0100	0100 101	1100	1100 110
0101	0101 010	1101	1101 001
0110	0110 011	1110	1110 000
0111	0111 100	1111	1111 111

This code has generator matrix and parity check matrix given by

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}, \qquad H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Note that the columns of H consist of the numbers one through seven, written in binary. This allows for very easy decoding. Note that a message string $x \in \mathbb{F}_2^4$ is encoded as a codeword $y = xG \in \mathbb{F}_2^7$. Suppose that this word is transmitted and a word $y' \in \mathbb{F}_2^7$ is received. The receiver first computes the syndrome $Hy^T \in \mathbb{F}_2^3$.

- If the syndrome is zero, then y' is a legitimate codeword. We assume no bit errors occurred during transmission (otherwise at least three bit errors occurred, which we deem unlikely). We recover $x \in \mathbb{F}_2^4$ as the first four bits of y'.
- If the syndrome is nonzero, then the syndrome is the binary representation of some $i \in \{1, 2, \ldots, 7\}$. First switch the i-th bit of y' (i.e. add 1 mod 2 in the i-th coordinate) to obtain the unique closest codeword to y'. We assume the resulting codeword is actually y; so as before, we recover $x \in \mathbb{F}_2^4$ as the first four bits of y'.

For example consider the message word $x = 1101$, which is encoded as $y = xG = 1101001$. Suppose this codeword is transmitted but that due to interference, the string $y' = 1111001$ is received. We compute the syndrome

$$Hy^T = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

This is the binary representation of 3, so switch the third bit of y' to recover $y = 1101001$ and thereby we find $x = 1101$.

Appendix A6: Invariant Theory

Let $R = \mathbb{C}[X_1, X_2, \ldots, X_n]$, the algebra of all polynomials in $X_1, X_2, \ldots, X_n$ with complex coefficients. For every invertible matrix $A = \left[a_{ij}\right]_{i,j} \in GL_n(\mathbb{C})$ and polynomial $f(X_1, \ldots, X_n) \in R$ we define

$$(Af)(X_1, \ldots, X_n) = f((X_1, \ldots, X_n)A) = f(\textstyle\sum_i a_{i1}X_i,\ \sum_i a_{i2}X_i,\ \ldots,\ \sum_i a_{in}X_i).$$

It is easy to see that $A(f+g) = Af + Ag$ and $A(fg) = (Af)(Ag)$. The map $f \mapsto Af$ is therefore an automorphism of the algebra R. Moreover $(AB)(f) = A(Bf)$ for all $A, B \in GL_n(\mathbb{C})$ since

$$\begin{aligned} A(Bf)(X_1, \ldots, X_n) &= (Bf)((X_1, \ldots, X_n)A) \\ &= f((X_1, \ldots, X_n)AB) \\ &= ((AB)f)(X_1, \ldots, X_n). \end{aligned}$$

This means that we have a (left) action of $GL_n(\mathbb{C})$ on R.

Let $\pi : G \to GL_n(\mathbb{C})$ be a complex linear representation of a group G. Recall that this simply means that π is a group homomorphism from G to $GL_n(\mathbb{C})$. Then G also acts on R, where the action of $g \in G$ on a polynomial f is given by

$$f \mapsto \pi(g)f.$$

The set of **invariants** of G (which of course depends on the choice of representation π) is the set of all polynomials fixed by every element of G:

$$R^G = \{f \in R : \pi(g)f = f \text{ for all } g \in G\}.$$

(Unfortunately this conflicts with the use of superscripts for G-orbits; but the notation is by now too standard to adjust.) Note that R^G is a subring of G; it is called the **ring of invariants** of G. The main question is to determine R^G. A first observation is that every $A \in GL_n(\mathbb{C})$ maps k-homogeneous polynomials to k-homogeneous polynomials, for each $k \geq 0$. Writing $R_k \subset R$ for the subspace of all k-homogeneous polynomials (with the zero polynomial included), then

$$R = \bigoplus_{k \geq 0} R_k$$

in which every summand is preserved by the action of G. By Section A3.2 we have

$$\text{(A6.1)} \qquad \dim R_k = \tbinom{n+k-1}{k}.$$

We obtain

$$R^G = \bigoplus_{k \geq 0} R_k^G$$

where R_k^G is the space of all k-homogeneous polynomials fixed by every element of G. Thus it suffices to determine each of the summands R_k^G.

A6.2 Example: The Orthogonal Group. Consider the real orthogonal group

$$G = \{A \in GL_n(\mathbb{R}) \,:\, AA^T = I\}$$

in its natural action on $\mathbb{C}^n$, i.e. we have simply the inclusion mapping $\pi : G \to GL_n(\mathbb{C})$, $A \mapsto A$. Since orthogonal transformations preserve the Euclidean norm, it is clear that G fixes the polynomial

$$\eta(X_1, X_2, \ldots, X_n) = X_1^2 + X_2^2 + \cdots + X_n^2 \,.$$

It follows that the ring of invariants R^G contains the subring generated by η, i.e.

$$R^G \supseteq \mathbb{C}[\eta] = \{f(X_1^2 + X_2^2 + \cdots + X_n^2) \,:\, f(T) \in \mathbb{C}[T]\}.$$

It turns out that equality holds: $R^G = \mathbb{C}[\eta]$.

A6.3 Example: The Symmetric Group. Consider the symmetric group $G = S_n$ in its natural permutation representation of degree n. Here every $\sigma \in S_n$ is represented by the corresponding permutation matrix, thus:

$$\pi(\sigma) = \left[\delta_{i^\sigma, j}\right]_{i,j} .$$

It is clear that the **elementary symmetric polynomials**

$$\begin{aligned}
e_1(X_1, X_2, \ldots, X_n) &= X_1 + X_2 + \cdots + X_n, \\
e_2(X_1, X_2, \ldots, X_n) &= \textstyle\sum_{1 \le i < j \le n} X_i X_j = X_1X_2 + X_1X_3 + \cdots + X_{n-1}X_n, \\
&\vdots \\
e_k(X_1, X_2, \ldots, X_n) &= \textstyle\sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} X_{i_1} X_{i_2} \cdots X_{i_k} \,, \\
&\vdots \\
e_n(X_1, X_2, \ldots, X_n) &= X_1X_2 \cdots X_n
\end{aligned}$$

are invariant under coordinate permutations. Therefore the ring of invariants contains the subring generated by $e_1, e_2, \ldots, e_n$, thus:

$$R^G \supseteq \mathbb{C}[e_1, e_2, \ldots, e_n].$$

Once again, although it is not obvious, equality holds: $R^G = \mathbb{C}[e_1, e_2, \ldots, e_n]$. For example, the polynomials

$$p_k(X_1, X_2, \ldots, X_n) = X_1^k + X_2^k + \cdots + X_n^k$$

are clearly invariants; according to our assertion, these may be expressed as polynomials in $e_1, e_2, \ldots, e_n$. Indeed, we find that

$$\begin{aligned} p_1 &= e_1; \\ p_2 &= e_1^2 - 2e_2; \\ p_3 &= e_1^3 - 3e_1e_2 + 3e_3; \end{aligned}$$

etc.

These examples are suggestive of the following general result: whenever G is finite, the ring of invariants is finitely generated. (The same conclusion holds more generally whenever the representation π is *unitary,* i.e. with respect to some basis of $\mathbb{C}^n$ the matrices $\pi(g)$ are unitary; in particular this happens whenever the subgroup $\pi(G) \subseteq GL_n(G)$ is a compact Lie group. The orthogonal group of Example A6.2 satisfies this condition.) The conclusion here is that there exists a finite list of polynomials $\eta_1, \eta_2, \ldots, \eta_m \in \mathbb{C}[X_1, \ldots, X_n]$ such that

$$R^G = \mathbb{C}[\eta_1, \eta_2, \ldots, \eta_m].$$

A list of such polynomials $\eta_k(X_1, \ldots, X_n)$ for $k = 1, 2, \ldots, m$, with m as small as possible, form what is called a **fundamental set of invariants**. Given G and π, algorithms are known (using Gröbner basis methods) for explicitly determining such a fundamental set of invariants. Such an algorithm is presented, for example, in [20], [60]. Although it is not our present purpose to describe such an algorithm here, nevertheless we outline one of the key ingredients: a criterion for testing whether a known list of invariants $\eta_1, \ldots, \eta_m$ is a fundamental set of invariants. A **graded subring** of R is a subring $S \subseteq R$ satisfying

$$S = \bigoplus_{k \geq 0} S_k$$

where $S_k = S \cap R_k$; for example that the ring of invariants of G is a graded subring. We define the **Hilbert series** of a graded subring $S \subseteq R$ by

$$H_S(t) = \sum_{k \geq 0} \big(\dim S_k\big) t^k.$$

A6.4 Theorem (Molien). Let $\pi : G \to GL_n(\mathbb{C})$ be a representation of a finite group G. Then the Hilbert series of the ring of invariants is given by

$$H_{R^G}(t) = \frac{1}{|G|} \sum_{g \in G} \frac{1}{\det(I - t\pi(g))}\,.$$

In order for Molien's Theorem to be useful, we must have some means of computing Hilbert series of some commonly arising subrings of R. Recall that polynomials $g_1, g_2, \ldots, g_m \in \mathbb{C}[X_1, X_2, \ldots, X_n]$ are **algebraically dependent** if there exists a nonzero polynomial $f(T_1, T_2, \ldots, T_m) \in \mathbb{C}[T_1, T_2, \ldots, T_m]$ such that $f(g_1, g_2, \ldots, g_m) = 0$. Otherwise, we say that $g_1, g_2, \ldots, g_m$ are **algebraically independent**.

A6.5 Theorem. Let $\eta_1, \eta_2, \ldots, \eta_m \in R$ be algebraically independent homogeneous polynomials of degree $d_1, d_2, \ldots, d_m$ respectively, and let $S = \mathbb{C}[\eta_1, \eta_2, \ldots, \eta_m] \subseteq R$, the subring generated by $\eta_1, \eta_2, \ldots, \eta_m$. Then the Hilbert series of S is given by

$$H_S(t) = \frac{1}{(1-t^{d_1})(1-t^{d_2})\cdots(1-t^{d_m})}\,.$$

Proof. The k-homogeneous component $S_k \subset S$ is spanned by $\{\eta_1^{e_1}\eta_2^{e_2}\cdots\eta_m^{e_m} \,:\, e_i \geq 0, \sum e_i = k\}$. In fact this is a basis for S_k: the polynomials $\eta_1^{e_1}\eta_2^{e_2}\cdots\eta_m^{e_m}$ cannot be linearly dependent, for otherwise the polynomials $\eta_1, \eta_2, \ldots, \eta_m$ would be algebraically dependent. So $\dim S_k$ is the number of m-tuples $(e_1, e_2, \ldots, e_m)$ of non-negative integers satisfying $e_1+e_2+\cdots+e_m = k$; but this is also the coefficient of t^k in

$$\frac{1}{(1-t^{d_1})(1-t^{d_2})\cdots(1-t^{d_m})}$$
$$= (1+t^{d_1}+t^{2d_1}+t^{3d_1}+\cdots)(1+t^{d_2}+t^{2d_2}+t^{3d_2}+\cdots)\cdots(1+t^{d_m}+t^{2d_m}+t^{3d_m}+\cdots). \qquad \square$$

A6.6 Example: The Trivial Group. Let $G = 1$ acting on $R = \mathbb{C}[X_1, \ldots, X_n]$, so that the ring of invariants is the full polynomial ring: $R^G = R$. By (A6.1) and the Binomial Theorem we obtain

$$H_{R^G}(t) = H_R(t) = \sum_{k\geq 0}(\dim R_k)t^k = \sum_{k\geq 0}\binom{n+k-1}{k}t^k = \frac{1}{(1-t)^n}\,.$$

This result agrees with the prediction of Molien's Theorem A6.4. Note that in this case a fundamental set of invariants is given by $X_1, X_2, \ldots, X_n$, each of degree 1, so that Theorem A6.5 gives the same result.

A6.7 Example: Even Polynomials. Let $n = 1$ and consider the group $G = \{1, g\}$ of order two, acting on $R = \mathbb{C}[X]$ by sign changes; here $\pi(g) = \big[-1\big]$. Clearly the ring of invariants is simply the subring of all even polynomials in X, i.e.

$$R^G = \mathbb{C}[X^2] = \{f(X^2) \,:\, f(T) \in \mathbb{C}[T]\}.$$

Clearly R_k^G is one-dimensional (spanned by X^k) whenever $k = 0, 2, 4, 6, \ldots$; and $R_k^G = 0$ whenever k is odd. Thus

$$H_{R^G}(t) = 1 + t^2 + t^4 + t^6 + \cdots = \frac{1}{1-t^2}\,.$$

Again since a fundamental set of invariants consists simply of $\eta(X) = X^2$ of degree 2, the same result is given by Theorem A6.5.

A6.8 Example: The Symmetric Group S_3. Consider the group $G = S_3$ acting naturally on $R = \mathbb{C}[X, Y, Z]$ by means of coordinate permutations. This is the special case of Example A6.3 with $n = 3$, where we claim that a fundamental set of invariants consists of the homogeneous polynomials

$$e_1(X,Y,Z) = X+Y+Z; \quad e_2(X,Y,Z) = XY+XZ+YZ \quad \text{and} \quad e_3(X,Y,Z) = XYZ.$$

We show how Molien's Theorem can be used to verify this result. The polynomials e_1, e_2 and e_3 are algebraically independent, which can be seen as follows. Suppose there exists a nonzero polynomial

$$h(T_1, T_2, T_3) = cT_1^{i_1}T_2^{i_2}T_3^{i_3} + \cdots \in \mathbb{C}[T_1, T_2, T_3]$$

such that $h(e_1, e_2, e_3) = 0$. Here $c \neq 0$ and $T_1^{i_1}T_2^{i_2}T_3^{i_3}$ is the lex-highest monomial appearing in h; that is, for every monomial $T_1^{j_1}T_2^{j_2}T_3^{j_3}$ appearing in h, either

$i_1 > j_1$; or
$i_1 = j_1$ and $i_2 > j_2$; or
$i_1 = j_1$ and $i_2 = j_2$ and $i_3 \geq j_3$.

Then the coefficient of $X^{i+j+k}Y^{j+k}Z^k$ in $h(e_1, e_2, e_3)$ is $c \neq 0$; in particular, $h(e_1, e_2, e_3)$ cannot vanish. By Theorem A6.5, the subring $S = \mathbb{C}[e_1, e_2, e_3] \subset R$ has Hilbert series

$$H_S(t) = \frac{1}{(1-t)(1-t^2)(1-t^3)}\,.$$

By Molien's Theorem, the Hilbert series of R^G is

$$H_{R^G}(t) = \frac{1}{6}\Big[\frac{1}{(1-t)^3} + \frac{3}{(1-t)(1-t^2)} + \frac{2}{1-t^3}\Big] = \frac{1}{(1-t)(1-t^2)(1-t^3)}\,.$$

We conclude that $S = R^G$ is the full ring of invariants. To see this, note that $S \subseteq R^G$ and so the k-homogeneous components satisfy $S_k \subseteq R_k^G$ for all $k \geq 0$; however $\dim S_k = \dim R_k^G$ by comparing coefficients of t^k in the corresponding Hilbert series; therefore equality must hold as claimed.

For further reading on the main themes of this course, the student is encouraged to consult additional sources as listed below:

Algebra: [49], [32]	Incidence geometry: [11]
Classical groups and their geometry: [61]	Projective geometry: [21], [12]
Coding Theory: [65], [14]	Projective planes: [31], [4]
Combinatorics: [13], [14]	Generalized quadrangles: [52]
Finite geometry: [25]	Generalized polygons: [63]

Bibliography

[1] R.W. Ahrens and G. Szekeres, 'On a combinatorial generalization of 27 lines associated with a cubic surface', *J. Austral. Math. Soc.* **10** (1969), 485–492.

[2] A. Balog and T.D. Wooley, 'Sums of two squares in short intervals', *Canad. J. Math.* **52** (2000), 673–694.

[3] A. Barlotti, 'Un'estensione del teorema di Segre-Kustaanheimo', *Bull. Un. Mat. Ital./* **10** (1955), 498–506.

[4] A. Beutelspacher, 'Projective planes', pp.107–136 in [11].

[5] A. Blokhuis, 'On the size of a blocking set in $PG(2,p)$', *Combinatorica* **14** (1994), 111–114.

[6] A. Blokhuis and G.E. Moorhouse, 'Some p-ranks related to orthogonal spaces', *J. Algebraic Combinatorics* **4** (1995), 295–316.
Available at `http://www.uwyo.edu/moorhouse/pub/orthog.pdf`

[7] A.E. Brouwer, 'A non-degenerate generalized quadrangle with lines of size four is finite', pp.47–49 in *Advances in Finite Geometries and Designs*, ed. J.W.P. Hirschfeld et. al., Oxford Univ. Press., Oxford, 1991.

[8] M. Brown, 'Ovoids of $PG(3,q)$, q even, with a conic section', *J. London Math. Soc.* **62** (2000), 569–582.

[9] R.H. Bruck and H.J. Ryser, 'The nonexistence of certain finite projective planes', *Canad. J. Math.***1** (1949), 88–93.

[10] A.A. Bruen, 'Blocking sets in finite projective planes', *SIAM J. Appl. Math.* **21** (1971), 380–392.

[11] F. Buekenhout, editor, *Handbook of Incidence Geometry: Buildings and Foundations,* Elsevier, Amsterdam, 1995.

[12] P.J. Cameron, *Projective and Polar Spaces*, QMW Maths Notes 13, London, 1991.
A second edition (2000) is available at `http://www.maths.qmul.ac.uk/~pjc/pps/`

[13] P.J. Cameron, *Combinatorics: Topics, Techniques, Algorithms,* Cambridge Univ. Press, Cambridge, 1994.

[14] P.J. Cameron and J.H. van Lint, *Designs, Graphs, Codes and their Links,* Camb. Univ. Press, Cambridge, 1991.

[15] G. Cherlin, 'Locally finite generalized quadrangles with at most five points per line', *Discrete Math.* **291** (2005), 73–79.

[16] S.S. Chern and P. Griffiths, 'Abel's theorem and webs', *Jahresberichte der Deut. Math. Ver.* **80** (1978), 13–110; also, 'Corrections and addenta to our paper: Abel's theorem and webs', same Journal, **83** (1981), 78–83.

[17] W. Cherowitzo, 'α-Flocks and Hyperovals', *Geom. Dedicata* **72** (1998), 221–246.

[18] W. Cherowitzo, 'Hyperoval Page', website.
`http://www-math.cudenver.edu/~wcherowi/research/hyperoval/hypero.html`

[19] J.H. Conway, P.B. Kleidman, and R.A. Wilson, 'New families of ovoids in O_8^+ ', *Geom. Dedicata* **6** (1988), 157–170.

[20] D.A. Cox, J.B. Little and D. O'Shea, *Ideals, Varieties, and Algorithms,* 3rd ed., Springer-Verlag, New York, 2007.

[21] H.S.M. Coxeter, *Projective Geometry,* 2nd ed., University of Toronto Press, 1974.

[22] T. Czerwinski and D. Oakden, 'The translation planes of order twenty-five', *J. Combin. Theory Ser. A* **59** (1992), 193–217.

[23] P. Deligne, 'La conjecture de Weil', *Publ. Math. IHES* **43** (1974), 273–307.

[24] P. Dembowski, 'Möbiusebenen gerader Ordnung', *Math. Ann.* **157** (1964), 179–205.

[25] P. Dembowski, *Finite Geometries,* Springer Verlag, Berlin, 1968.

[26] U. Dempwolff, 'Translation planes of order 27', *Des. Codes and Crypt.* **27** (1994), 105–121.

[27] W. Feit and G. Higman, 'The nonexistence of certain generalized polygons', *J. Algebra* **1** (1964), 114–131.

[28] A. Gunawardena and G.E. Moorhouse, 'The nonexistence of ovoids in $O_9(q)$', *Europ. J. Combinatorics* **18** (1997), 171–173.
Available at `http://www.uwyo.edu/moorhouse/pub/9dim.pdf`

[29] D.R. Hughes, 'Generalized incidence matrices over group algebras', *Ill. J. Math.* **1** (1957), 545–551.

[30] D.R. Hughes, 'Collineations and generalized incidence matrices', *Trans. Amer. Math. Soc.* **86** (1957), 284–296.

[31] D.R. Hughes and F.C. Piper, *Projective Planes,* Springer Verlag, New York, 1973.

[32] T.W. Hungerford, *Algebra*, Springer-Verlag, New York, 1974.

[33] W.M. Kantor, 'Ovoids and translation planes', *Canad. J. Math.* **34** (1982), 1195–1207.

[34] N.M. Katz, 'An overview of Deligne's proof of the Riemann hypothesis for varieties over finite fields', *Proc. Symp. Pure Math.* **28**, Amer. Math. Soc., Providence, R.I., 1976, pp.275–305.

[35] N. Koblitz, *Introduction to Elliptic Curves and Modular Forms,* 2nd ed., Springer-Verlag, New York, 1993.

[36] C.W.H. Lam, 'The search for a finite projective plane of order 10', *Amer. Math. Monthly* **98** (1991), 305–318.

[37] C.W.H. Lam, G. Kolesova and L. Thiel, 'A computer search for finite projective planes of order 9', *Discrete Math.* **92** (1991), 187–195.

[38] C.W.H. Lam, L. Thiel and S. Swiercz, 'The nonexistence of finite projective planes of order 10', *Canad. J. Math.* **41** (1988), 1117-1123.

[39] E. Landau, *Handbuch der Lehre der Verteilung der Primzahlen, Bd. 2,* Teubner, Leipzig-Berlin, 1909.

[40] S. Lie, 'Bestimmung aller Flächen, die in mehrfacher Weise durch Translationsbewegung einer Kurve erzeugt werden', *Arch. für Math.* Bd. 7, Heft 2 (1882), 155–176.

[41] L. Lunelli and M. Sce, 'K-archi completi nei piani proietivi desarguesiani di rango 8 e 16', Centro Calcoli Numerici, Politecnico di Milano, 1958.

[42] G.E. Moorhouse, 'Bruck nets, codes, and characters of loops', *Designs, Codes and Cryptography* **1** (1991), 7–29.
Available at `http://www.uwyo.edu/moorhouse/pub/loop1.pdf`

[43] G.E. Moorhouse, 'Ovoids from the E_8 root lattice', *Geom. Dedicata* **46** (1993), 287–297.
Available at `http://www.uwyo.edu/moorhouse/pub/e8.pdf`

[44] G.E. Moorhouse, 'Ovoids and translation planes from lattices', in *Mostly Finite Geometries,* ed. N. L. Johnson, Marcel Dekker Inc., 1997, 127–134.
Available at `http://www.uwyo.edu/moorhouse/pub/iowa1.pdf`

[45] G.E. Moorhouse, 'Ranks of nets', *Quasigroups and Related Systems* **14** (2006), 61–72.
Available at `http://www.uwyo.edu/moorhouse/pub/qrs.pdf`

[46] G.E. Moorhouse, 'Projective planes of small order', technical report.
Available at `http://www.uwyo.edu/moorhouse/pub/planes/`

[47] G.E. Moorhouse, 'Generalised polygons of small order', technical report.
Available at `http://www.uwyo.edu/moorhouse/pub/genpoly/`

[48] G.E. Moorhouse, 'On projective planes of order less than 32', in *Finite Geometries, Groups, and Computation,* ed. A. Hulpke et. al., de Gruyter, Berlin, 2006, pp.149–162.
Available at `http://www.uwyo.edu/moorhouse/pub/wyoming.pdf`

[49] G.E. Moorhouse, *Abstract Algebra I, Lecture Notes for Math 5550,* University of Wyoming, revised 2003.
Available at `http://www.uwyo.edu/moorhouse/handouts/algebra.pdf`

[50] J. di Paola, 'On minimum blocking coalitions in small projective plane games', *SIAM J. Appl. Math.* **17** (1969), 378–392.

[51] S.E. Payne, 'Nonisomorphic generalized quadrangles', *J. Algebra* **18** (1971), 201–212.

[52] S.E. Payne, 'A new infinite family of generalized quadrangles', *Congressus Numerantium* **49** (1985), 115–128.

[53] S.E. Payne and J.A. Thas, *Finite Generalized Quadrangles*, Pitman, Boston, 1984.

[54] T. Penttila, G.F. Royle and M.K. Simpson, 'Hyperovals in the known projective planes of order 16', *J. Combinatorial Designs* **4** (1998), 59–65.

[55] B. Qvist, 'Some remarks concerning curves of the second degree in a finite plane', *Ann. Acad. Sci. Fenn.*, no. 134 (1952), pp.1–27.

[56] G. Royle, 'Projective planes of order 16', technical report.
Available at `http://people.csse.uwa.edu.au/gordon/remote/planes16/`

[57] R. Scharlau, 'Buildings', pp.477–645 in [11].

[58] J.-P. Serre, *A Course in Arithmetic*, Springer Verlag, New York, 1973.

[59] E.E. Shult and J.A. Thas, 'm-Systems of polar spaces', *J. Combin. Theory Ser. A* **68** (1994), 184–204.

[60] B. Sturmfels, *Algorithms in Invariant Theory*, Springer-Verlag, Vienna, 1993.

[61] D.E. Taylor, *The Geometry of the Classical Groups*, Heldermann Verlag, Berlin, 1992.

[62] J.A. Thas, 'Ovoids and spreads of finite classical polar spaces', *Geometriae Dedicata* **10** (1981), 135–144.

[63] J.A. Thas, 'Generalized polygons', pp.383–431 in [11].

[64] J. Tits, *Buildings of Spherical Type and finite BN-pairs*, Springer-Verlag, New York, 1974.

[65] J.H. van Lint, *Introduction to Coding Theory*, 2nd ed., Springer-Verlag, Berlin, 1992.

[66] H.A. Wilbrink, 'A note on planar difference sets', *J. Comb. Theory Ser. A* **38** (1985), 94–95.

Index